



Sónia Catarina Martins Roxo

Licenciada

**Integração de informação secundária
na modelação geoestatística da
qualidade de solos em locais
potencialmente contaminados.
Aplicação à área subjacente de uma
antiga refinaria.**

Dissertação para obtenção do Grau de Mestre em
Engenharia Geológica (Georrecursos)

Orientador: José António de Almeida, Prof. Auxiliar –
FCT/UNL

Júri:

Presidente: Prof. Doutor(a) Maria da Graça Azevedo de Brito
Vogal(ais): Prof. Doutor(a) Maria Teresa Condesso de Melo
Prof. Doutor(a) José António de Almeida

**Integração de informação secundária na modelação
geoestatística da qualidade de solos em locais
potencialmente contaminados. Aplicação à área
subjacente de uma antiga refinaria.**

Copyright em nome de Sónia Catarina Martins Roxo, da FCT/UNL e da UNL

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa tem o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

AGRADECIMENTOS

Este estudo não é apenas resultado de um empenho individual, mas sim de um conjunto de esforços que o tornaram possível e sem os quais teria sido muito mais difícil chegar ao fim desta etapa, que representa um importante marco na minha vida pessoal e profissional. Desta forma, manifesto a minha gratidão a todos os que estiveram presentes nos momentos de dúvidas, angústia, de insegurança e de satisfação.

Ao meu orientador, Prof. Doutor José António Almeida, pela oportunidade e encorajamento na abordagem deste tema, pela sua excelente capacidade de orientação científica, pelo empenho e disponibilidade sempre manifestada em ajudar, pelo entusiasmo e motivação.

Aos meus pais e marido, que sempre acreditaram no meu empenho, pela amizade, motivação e carinho. Ao meu filhote, pelo tempo que tive de abdicar dele, durante os primeiros meses de vida. Por me darem força para seguir em frente em mais esta etapa académica e por não me deixarem olhar para trás e desistir.

RESUMO

Este trabalho apresenta uma metodologia para caracterizar espacialmente a contaminação em hidrocarbonetos totais do petróleo (Total Petroleum Hydrocarbons - TPH) no local de um antigo parque de combustíveis. A metodologia proposta foi desenvolvida com algoritmos de estimação e simulação geoestatística 3D e integra informação principal e secundária.

Foram utilizados dados de 131 sondagens de uma anterior campanha de prospecção, para uma área de aproximadamente 14,5ha. As variáveis secundárias são apresentadas por classes (variável categórica) relativamente a três análises expeditas macroscópicas e sensoriais ao solo recolhido das sondagens. Ao todo foram analisados e codificados 820 troços, que constituem o comprimento total das sondagens. Dos 820 troços, 276 foram enviados para análise laboratorial tendo sido reportados os teores em TPH (informação principal). Comparativamente aos dados laboratoriais, os dados das análises sensoriais são muito expeditos e, por isso, mais abundantes; todavia têm um elevado grau de subjectividade na discriminação da contaminação. Pelo contrário, as determinações analíticas são menos expeditas e muito mais caras, mas os resultados são mais fiáveis do ponto de vista de identificar a contaminação.

A metodologia proposta desenvolve-se em três etapas. Em primeiro lugar, procedeu-se à análise estatística preliminar para determinar qual das análises sensoriais e respectivas classes (cor, cheiro e reacção ao óleo) seriam mais adequadas para discriminar os teores em TPH. De seguida, fez-se a estimação 3D das classes da variável seleccionada por krigagem da indicatriz. Finalmente, na terceira etapa, faz-se a simulação dos teores em TPH. Na simulação utilizou-se uma variante inovadora do algoritmo de simulação sequencial directa com condicionamento a histogramas e médias locais. A metodologia proposta foi validada com um teste de validação cruzada, construído a partir de um subconjunto aleatório de 20% das amostras.

Palavras-Chave:

Mapa de contaminação de solos; hidrocarbonetos totais do petróleo; informação principal e secundária; krigagem da indicatriz; simulação sequencial directa.

ABSTRACT

This work presents a methodology to map the TPH (Total Petroleum Hydrocarbons) content in a deactivated oil storage site. The proposed methodology was developed with algorithms for 3D geostatistical estimation and simulation and integrates soft and hard data.

Data from 131 boreholes of a previous survey were used for an area of approximately 14,5ha. Soft data derives from three expedite macroscopic and sensory analyses to the soil collected from the surveys. Altogether 820 soil samples were analyzed and coded by classes (categorical variable) making up the total length of the boreholes. Of the 820 samples, 276 were analysed in TPH contents (hard data). Compared to laboratory data, data from sensory analysis are much quicker and therefore more plentiful, yet have a high degree of subjectivity in the discrimination of contamination. Rather, the analytical determinations are less resourceful and more expensive, but the results are more reliable in terms of identifying the TPH contents.

The proposed methodology develops in three main steps. First, a preliminary statistical analysis was conducted to check the relationship between the soft data (colour, smell and water-oil reaction) and the hard data and select the best soft variable. Then, soft variable was estimated for the entire volume by 3D indicator kriging. Finally, TPH contents were conditionally simulated. For simulation, an innovative variant of the direct sequential simulation with conditioning to local histograms and local means were used. Validation was done through a cross-validation test build with a random sample of 20% of the boreholes.

Key-words:

Soil contamination mapping; Total Petroleum Hydrocarbons (TPH); hard and soft data; indicator kriging; direct sequential simulation.

ÍNDICE GERAL

1. Introdução	1
1.1. Apresentação do estudo, sua importância e objectivos	1
1.2. Organização da tese	4
1.3. Caracterização Sumária da Área em Estudo e Amostragem do Local	4
1.4. Enquadramento Geológico e Hidrogeológico da Área de Estudo	6
2. Metodologia e Fundamentos Teóricos	9
2.1. Metodologia.....	9
2.2. Fundamentos teóricos geoestatísticos	13
2.2.1. Variáveis Aleatórias.....	14
2.2.2. Variografia e ajustamento de modelos teóricos	15
2.2.3. Estimação por krigagem.....	17
2.2.4. Formalismo da indicatriz.....	19
2.2.4.1. Populações multifásicas com p classes.....	19
2.2.4.2. Variografia de populações multifásicas com p classes.....	20
2.2.5. Estimação da Variável Indicatriz Multifásica	21
2.2.6. Simulação Sequencial Directa com Condicionamento a Histogramas Locais.....	21
2.3. Testes de Validação Cruzada.....	23
3. Análise Estatística Preliminar	25
4. Modelação Geoestatística dos Teores em TPH	35
4.1. Modelação das Classes de Reacção ao Óleo	35
4.1.1. Variogramas multifásicos e ajustamento de modelos teóricos	35
4.1.2. Krigagem multifásica das classes de reacção ao óleo	37
4.2. Co-estimação dos Teores em TPH Condicionado às Classes de Reacção ao Óleo, Fase de Teste de validação cruzada	43
4.2.1. Cálculo de variogramas e ajustamento de modelos teóricos.....	43
4.2.2. Teste de validação cruzada e análise de resultados	44
4.3. Co-simulação de Imagens dos teores em TPH Condicionado às Classes de Reacção ao Óleo.....	50
4.3.1. Apresentação e Validação de Resultados	50
4.3.2. Binarização das Imagens dos Teores em TPH.....	57
5. Conclusões	61
6. Bibliografia.....	63

ÍNDICE DE FIGURAS

Figura 2.1 – Diagrama da metodologia proposta para a caracterização da contaminação em TPH na área da antiga refinaria em estudo.	12
Figura 3.1 – Planta de localização das sondagens e piezómetros e discretização da área de estudo numa malha de blocos de 5 x 5 x 0,5 metros.....	25
Figura 3.2 – Histograma de profundidade máxima atingida pelas sondagens.....	26
Figura 3.3 – Histograma dos comprimentos dos suportes amostrados em termos de classe sensorial e análises químicas.	27
Figura 3.4 – Histograma do número de amostras por classe e tipo de análise sensorial.	28
Figura 3.5 – Histograma cumulativo condicional dos teores em TPH por classe de cor.....	29
Figura 3.6 – Histograma cumulativo condicional dos teores em TPH por classe de cheiro.....	29
Figura 3.7 – Histograma cumulativo condicional dos teores em TPH por classe de reacção ao óleo.....	30
Figura 3.8 – Histograma cumulativo condicional dos teores em TPH por classe de reacção ao óleo após agregação das classes 2+3+4.....	31
Figura 3.9 – Histograma cumulativo condicional dos teores em TPH por classe de reacção ao óleo após agregação das classes 0+1 e 2+3+4.	32
Figura 4.1 – Variogramas das classes de reacção ao óleo (classes 0, 1 e 2+3+4) na direcção omnidireccional horizontal.....	36
Figura 4.2 – Variogramas das classes de reacção ao óleo (classes 0, 1 e 2+3+4) na direcção vertical	36
Figura 4.3 – Variogramas experimentais multifásicos das classes de reacção ao óleo nas direcções horizontal e vertical e ajustamento com modelo exponencial.	36
Figura 4.4 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 1 [0,00m; 0,50m].	37
Figura 4.5 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 2 [0,50m; 1,00m].	38
Figura 4.6 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 3 [1,00m; 1,50m].	38
Figura 4.7 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 4 [1,50m; 2,00m].	39
Figura 4.8 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 5 [2,00m; 2,50m].	39

Figura 4.9 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 6 [2,50m; 3,00m].	40
Figura 4.10 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 7 [3,00m; 3,50m].....	40
Figura 4.11 – Classes de reacção ao óleo (0+1 e 2+3+4) para os 7 horizontes estimados.	41
Figura 4.12 – Classes de reacção ao óleo (0, 1, 2+3+4) para os 7 horizontes estimados.....	42
Figura 4.13 – Variogramas experimentais dos teores em TPH nas direcções horizontal e vertical e ajustamento com um modelo esférico.	43
Figura 4.14 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base SSD_MG. ..	45
Figura 4.15 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base SSD_ML_CRO (01_234).....	45
Figura 4.16 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base SSD_ML_CRO (0_1_234).....	46
Figura 4.17 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base SSD_HL_ML_CRO (01_234).....	46
Figura 4.18 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base SSD_HL_ML_CRO (0_1_234).....	47
Figura 4.19 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base SSD_HL_ML_CRO&Z (01_234).	47
Figura 4.20 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base SSD_HL_ML_CRO&Z (0_1_234).	48
Figura 4.21 – Três imagens simuladas de TPH para o horizonte (0-0,5m).	51
Figura 4.22 – Três imagens simuladas de TPH para o horizonte (0,5-1m).	51
Figura 4.23 – Três imagens simuladas de TPH para o horizonte (1-1,5m).	51
Figura 4.24 – Três imagens simuladas de TPH para o horizonte (1,5-2m).	52
Figura 4.25 – Três imagens simuladas de TPH para o horizonte (2-2,5m).	52
Figura 4.26 – Três imagens simuladas de TPH para o horizonte (2,5-3m).	52

Figura 4.27 – Três imagens simuladas de TPH para o horizonte (3-3,5m).	53
Figura 4.28 – Variogramas de uma imagem simulada para as 3 direcções ortogonais: NS, EW e vertical e sobreposição com o modelo teórico ajustado aos dados experimentais.	53
Figura 4.29 – Valores médios locais em TPH por horizonte calculados com base nas 100 simulações.....	54
Figura 4.30 – Variâncias locais em TPH por horizonte calculados com base nas 100 simulações.....	55
Figura 4.31 –Histogramas cumulativos condicionais de TPH por classe de reacção ao óleo para a realização #1.	56
Figura 4.32 – Imagem de probabilidade local dos teores em TPH excederem o valor de corte de 1000 ppm para os 7 horizontes estimados.	58
Figura 4.33 – Imagem binária de teores de TPH acima e abaixo de 1000 ppm, respectivamente categorias 1 e 0, para os 7 horizontes estimados.....	59

ÍNDICE DE TABELAS

Tabela 3.1 – Frequências absolutas cruzadas entre as classes de reacção ao óleo e as classes de cor e cheiro.....	28
Tabela 3.2 – Estatísticos básicos condicionais de TPH (ppm) por classe de análise sensorial.	30
Tabela 3.3 – Posicionamento da malha de blocos para o modelo geoestatístico.....	32
Tabela 3.4 – Extracto do ficheiro de dados para uma sondagem.	33
Tabela 3.5 – Estatísticos básicos das classes de reacção ao óleo e teores em TPH (ppm) por classe de análise sensorial até à profundidade máxima de 3,5 metros.	33
Tabela 3.6 – Estatísticos básicos das classes de reacção ao óleo e teores em TPH (ppm) por classe de análise sensorial e por intervalo de profundidade.	34
Tabela 4.1 – Proporções de cada classe por horizonte e proporções globais (imagem estimada / dados experimentais).	43
Tabela 4.2 – Análise estatística comparativa da diferença entre o valor verdadeiro e os 100 valores simulados amostra a amostra.	49

1. INTRODUÇÃO

1.1. APRESENTAÇÃO DO ESTUDO, SUA IMPORTÂNCIA E OBJECTIVOS

As questões relativas à contaminação de solos e de recursos hídricos, quer superficiais quer subterrâneos, são uma preocupação actual, uma vez que os recursos solos e água são cada vez mais escassos. Apesar da água subterrânea ser considerada actualmente um recurso natural estratégico em Portugal, continuam a ser colocados, com frequência, sob risco devido a actividades antrópicas. Diversos tipos de contaminação têm contribuído para a deterioração gradual da qualidade da água subterrânea, sem que muitas vezes tenham existido acções legais efectivas para a proteger.

Considera-se um solo contaminado, quando as suas características naturais se encontram modificadas devido à adição de compostos químicos, o que limita o seu uso e provoca a degradação da qualidade das águas, constituindo um risco para a saúde pública e para o meio ambiente (Brito, 2005).

Até recentemente, a problemática da contaminação dos solos prendia-se principalmente com a sua utilização na agricultura, como fonte de produção alimentar e subsistência. Os solos em meio urbano encontram-se naturalmente mais vulneráveis do que os solos em meio rural, uma vez que no meio urbano existe uma maior pressão derivada da actividade humana, construções desenfreada e, de todas as actividades industriais, que originam naturalmente uma maior contaminação dos solos e águas (Pereira *et al*, 1997,2001; Brito *et al*, 2006; Silva e Almeida, 2009).

O crescente crescimento e desenvolvimento populacional e urbano, obriga à necessidade de reconversão e consequente reocupação dos terrenos, seja para fins urbanos, industriais ou agrícolas. Este factor obriga muitas vezes a que seja necessário averiguar, mediante o uso anterior desses terrenos, o grau de contaminação destes.

A diversidade de indústrias, muitas vezes existentes num local, origina uma grande diversidade de materiais e substâncias que provocam alterações nas propriedades químicas dos solos, originando diferentes graus de contaminação. Deste modo a adição de substâncias poluentes nos solos, para além de alterar a natureza química destes, tem efeitos adversos no tipo de construções, materiais de construção a utilizar, no meio ambiente (flora e fauna locais), actividades agrícolas, uma vez que a acumulação de substâncias tóxicas nos vegetais torna-os impróprios para consumo, para além de constituir um factor de elevada perigosidade para a saúde pública.

Consoante o tipo e a natureza do terreno, bem como o uso que lhe foi dado e do uso que se pretende dar na ocupação futura, assim depende o tipo de metodologias e estratégias de amostragem a utilizar para uma correcta e precisa avaliação da extensão do estado de contaminação de um determinado local (Brito, 2005).

Na realidade, um local contaminado, não é mais que um local onde a ocorrência de concentrações de substâncias químicas excedem os valores de referência constantes nas normas adaptadas em cada país. No caso de Portugal, segundo a Agência Portuguesa do Ambiente (APA), as normas são as constantes no "*Guidelines for use at contaminated sites in Ontario*" (OMEE, 2011).

Nos últimos anos têm sido feitos vários esforços no sentido de requalificar várias áreas ambientalmente degradadas devido sobretudo à contaminação dos solos, são disso exemplo a zona da Expo 98, em Lisboa; a requalificação de uma unidade fabril da Portucel Recicla, em Mourão; a descontaminação de solos e águas subterrâneas da fábrica Gist-Brocades, em Matosinhos; o complexo químico de Estarreja; a descontaminação da zona 1 do parque empresarial do Barreiro – Quimiparque; descontaminação dos solos da ex-Siderurgia Nacional; descontaminação dos solos na antiga Mina de São Domingos, em Mértola; descontaminação de solos em antigas refinarias da Petrolgal, entre elas a área de estudo.

É necessário definir critérios, com base em análises de risco, de modo a que o dinheiro gasto na requalificação/remediação de áreas contaminadas seja feita de forma correcta.

Neste sentido, a requalificação/remediação de um local contaminado deverá proteger a saúde pública, bem como o meio ambiente e, reabilitar o local afectado de modo a que o futuro uso do solo apresente características adequadas. Assim, as acções de descontaminação devem ter como objectivo a minimização do risco para o receptor.

Actualmente as tecnologias aplicadas no tratamento de terrenos caracterizam-se pela utilização generalizada de técnicas convencionais de engenharia civil e geotecnia, a técnica "dig & dump", que no fundo não é mais que a escavação e transporte para aterro; isolamento dos solos contaminados in situ, seja por confinamento lateral (cortinas de estacas-prancha, paredes moldadas, entre outras), por confinamento de topo (selagem com betão, betuminoso, entre outras) ou por confinamento total (encapsulamento com injeção de caldas impermeabilizantes, resinas, entre outras); ou tratamento dos solos contaminados in situ, seja através da solidificação (aplicação de aditivos que permitam aumentar a resistência, diminuir a permeabilidade, entre outros), da estabilização (aplicação de aditivos reactivos que permitam diminuir a toxicidade e a mobilidade dos agentes contaminantes), da extracção através da injeção de ar, ar + CH₄, calor, entre outros, no caso do contaminante se localizar abaixo do nível freático através da técnica de extracção por bombagem (*pump & treat*) e injeção (*air sparging*), através da biorremediação (biodegradação, transformação e bioacumulação por acção de microorganismos, fitorremediação, entre outros) ou através da adição de reagentes químicos de modo a acelerar os processos de oxidação-redução, imobilização, lixiviação, entre outros (Brito, 2005).

Qualquer processo de requalificação/ remediação de um local contaminado tem um custo elevado, é um processo moroso e apresenta uma baixa relação custo-benefício. Não existe um método de tratamento de uma área contaminada específica que seja melhor que outro. Cada caso deve ser analisado individualmente, verificando e analisando as suas singularidades.

Assim, de modo a minimizar os custos, devem ser efectuados trabalhos de investigação e amostragem, como determinações expeditas no local, com análises laboratoriais, tratamento de dados com recurso a *softwares* de modelação e análise de risco do local em estudo.

De modo a realizar uma análise de risco e de remediação/recuperação, é necessário dispor de mapas que permitam caracterizar o estado de contaminação de um determinado local. Este tipo de mapas, os quais apresentam os quantitativos totais e a análise de risco associada, permitem comparar várias opções de remediação/recuperação de modo a otimizar custos.

Quando ocorre um derrame de produtos refinados do petróleo, estes sofrem vários processos de homogeneização, nomeadamente a adsorção de hidrocarbonetos pelas partículas do solo, a volatilização de hidrocarbonetos e a dissolução destes na água. Importa referir que menos de 5% dos produtos refinados do petróleo se dissolvem na água. Destes, os hidrocarbonetos aromáticos são os produtos derivados do petróleo mais solúveis, dos quais se destaca o benzeno (C₆H₆). Quando ocorre um derrame de produtos refinados do petróleo, o benzeno é geralmente o principal contaminante, uma vez que apresenta uma elevada toxicidade e mobilidade, quando comparado com outros hidrocarbonetos. Os hidrocarbonetos aromáticos policíclicos (HAP) são na sua maioria classificados como componentes orgânicos semi-voláteis, devido à sua baixa volatilidade. Geralmente os HPA não se dissolvem facilmente na água, devido à sua baixa solubilidade e elevada mobilidade, agregando-se por isso, aos sedimentos e solos.

A contaminação por hidrocarbonetos de um determinado local, pode ser, de forma expedita, avaliada por um técnico experiente, que através da análise da cor, cheiro e a reacção ao óleo (cor e intensidade da película de hidrocarbonetos que ascende à superfície da água) consegue determinar o grau de contaminação, que posteriormente é validado através de análises químicas em laboratório.

Assim, o trabalho de campo, deverá ter por objectivo, a recolha do maior número possível de dados sensoriais e analisar uma fracção destes em laboratório. As amostras que são enviadas para análise laboratorial devem corresponder sempre aos locais onde se obtiveram os dados sensoriais, para que possa haver uma validação das amostras.

O objectivo do presente estudo é apresentar e testar uma metodologia que permita caracterizar espacialmente a contaminação de solos por hidrocarbonetos (TPH) de um antigo parque de combustíveis, através da análise de dados *hard* e *soft* (informação primária e secundária, respectivamente).

Este estudo foi enquadrado no projecto “CRUDE – Desenvolvimento de novas estratégias de amostragem, análise e modelação para caracterização da contaminação dos solos e águas subterrâneas por contaminantes orgânicos”, financiado pela Fundação para a Ciência e Tecnologia PTDC/CTE-GEX/72959/2006.

1.2. ORGANIZAÇÃO DA TESE

O presente trabalho desenvolve-se em 5 capítulos. No primeiro faz-se a apresentação do problema das contaminações e a necessidade dos estudos.

No segundo capítulo apresentam-se a metodologia e os fundamentos teóricos necessários à compreensão do presente estudo.

No capítulo 3 faz-se uma análise estatística detalhada da informação utilizada no presente estudo do ponto de vista uni e bivariado.

No capítulo 4 apresentam-se e comentam-se os resultados do caso de estudo, e no capítulo 5 fazem-se os comentários finais.

1.3. CARACTERIZAÇÃO SUMÁRIA DA ÁREA EM ESTUDO E AMOSTRAGEM DO LOCAL

Importa referir, em primeiro lugar, que por efeitos de confidencialidade não é possível dar todos os detalhes acerca do local em estudo, e cujos dados foram utilizados neste trabalho para a apresentação de uma metodologia de caracterização de locais contaminados por hidrocarbonetos. Também não houve intervenção nem acompanhamento na execução dos trabalhos de amostragem.

A área objecto de estudo neste trabalho é um antigo parque de combustíveis, entretanto desactivado, que durante o seu período de actividade armazenou produtos refinados do petróleo (gasolinas e gasóleo) e petróleo bruto em tanques. Considerando a possibilidade do desmantelamento do parque, e futura reutilização da área para fins industriais ou urbanos, foi realizada uma amostragem aos solos e água subterrânea com vista à caracterização espacial da contaminação dos solos e avaliação da qualidade das águas subterrâneas. Importa referir que o presente trabalho apenas recai sobre a contaminação dos solos.

A área de estudo tem cerca de 14,5ha e localiza-se nas proximidades do rio Tejo. Inclui várias infra-estruturas de armazenamento e transporte, de que se destacam cinco bacias de segurança dos tanques com cerca de 11000 metros quadrados cada, um sistema de tubagens para transporte de combustíveis (*pipeline*), uma estação de tratamento de águas residuais e um sistema de incêndio e refrigeração. Na altura em que foi realizada a campanha de amostragem, as infra-estruturas ainda se encontravam todas no local, mas os produtos armazenados já tinham sido removidos.

As bacias de segurança eram drenadas através de canais de cimento abertos. A água e os resíduos de combustíveis escoavam através de um sistema de esgoto subterrâneo até à estação de tratamento de águas residuais localizada no centro do parque. A água depois de

tratada no separador óleo/água era descarregada numa pequena linha de água que desagua no rio Tejo.

A estratégia de amostragem foi fazer furos no solo até ao substrato (máximo de 8 metros), permitindo a delimitação vertical de zonas contaminadas. Ao todo foram utilizados para este estudo dados de 131 sondagens. A prospecção incidiu sobre toda a área do parque, com particular destaque para as bacias de drenagem dos tanques, a estação de tratamento de águas residuais, zonas de armazenagem de resíduos e todas as áreas indiciadas como potencialmente contaminadas. O plano de prospecção foi ajustado às infra-estruturas existentes e às observações de campo que iam sendo obtidas. O plano de amostragem incluiu também a instalação de poços de monitorização da qualidade das águas subterrâneas, cujos dados não foram considerados neste trabalho.

Todos os furos foram realizados por uma sonda com recuperação de testemunho (amostra não destrutiva) cuja força foi aplicada por um martelo eléctrico de percussão. Este tipo de sonda permite recuperar amostras de camadas específicas e obter perfis detalhados dos furos.

Tendo em conta enquadramento geológico e pedológico do local, assim como o tipo e extensão de eventual contaminação, a metodologia de recolha de informação adoptada no local foi do tipo “Quick Scan Soil Survey”. Esta metodologia combina a recolha de observações sensoriais (cor, odor e reacção do óleo) e de amostras para análises químicas. Ao combinar estes dois tipos de informação, o número de amostras enviadas para análise química é menor, pelo que é mais económico.

A observação da reacção do óleo é um procedimento patenteado pela Arcadis Heidemij®. Esta observação sensorial consiste em detectar a presença de petróleo e seus componentes através da observação visual. Para tal utiliza-se uma tina com água limpa e onde se deposita uma pequena quantidade de solo. Os componentes do petróleo produzem um filme colorido na superfície da água, onde a intensidade do brilho e a cor deste filme dão uma indicação da concentração de TPH (componentes não voláteis), numa escala entre 0 e 5, sendo zero o menor indício e 5 o máximo.

Foram analisadas 778 amostras sensoriais, tendo sido analisadas em laboratório 159 amostras, de modo que as análises sensoriais fossem validadas com as análises laboratoriais. O número de análises sensoriais é significativamente superior às análises laboratoriais, devido a ser um método expedito e mais económico.

As amostras de solo enviadas para análise laboratorial foram escolhidas de acordo com as observações sensoriais, particularmente a reacção do óleo. As amostras de solo foram conservadas em frascos com tampas especiais com cobertura de *teflon* para evitar a evaporação dos componentes voláteis, e foram armazenadas e transportadas em caixas isotérmicas, por avião, para o laboratório da Alcontrol na Holanda.

O sucesso da mistura de informação sensorial com determinações analíticas depende do sucesso da captura por parte das análises sensoriais de indícios de hidrocarbonetos. Do ponto

de vista estatístico, é necessário que exista discriminação positiva nos valores obtidos (entre as classes sensoriais e os valores das análises na mesma amostra) pelo que foram observados os seguintes critérios na escolha das amostras a enviar para laboratório:

- Algumas amostras de solo com pouca ou nenhuma reacção do óleo – seleccionadas apenas para calibração das observações sensoriais das amostras classificadas como potencialmente não contaminadas (classes 0 e 1);
- Muitas amostras das áreas potencialmente contaminadas – estas amostras foram seleccionadas de modo a obter o nível máximo de contaminação do solo em TPH;
- Em alguns furos foram seleccionadas várias amostras seguidas para obter o gradiente vertical da contaminação e permitir a estimação de variogramas verticais;
- Em alguns furos com amostras indiciadas como contaminadas, foram seleccionadas amostras das camadas limpas para delimitar verticalmente as eventuais manchas de contaminação.

1.4. ENQUADRAMENTO GEOLÓGICO E HIDROGEOLÓGICO DA ÁREA DE ESTUDO

A área de estudo localiza-se na planície aluvionar da margem direita do rio Tejo, imediatamente a norte de uma linha de água.

Geologicamente a área de estudo localiza-se na Orla Meso-Cenozóica Ocidental, na Bacia do Tejo-Sado, formada por depósitos paleogénicos, miocénicos e pliocénicos recobertos em grande parte por depósitos quaternários.

A área de estudo é abrangida pela folha 34B – Loures, da Carta Geológica de Portugal à escala 1/50 000 (DGMSG, 1981). Descrevem-se de seguida com base na referida Carta, as unidades litostratigráficas aflorantes na área de estudo.

Quaternário

Holocénico

- **a – Aluviões** – Na área de estudo afloram as aluviões associados à margem direita do Rio Tejo, podendo atingir espessuras da ordem dos 60m. Correspondem nesta zona a um complexo fluvio-marinho constituído por lodos com intercalações de areias mais ou menos lodosas, com cores acinzentadas escuras. Na base das aluviões é também frequente a ocorrência de níveis de cascalheiras de calhaus rolados.

Plistocénico

- **Q – Depósitos de Terraços Fluviais** – Os depósitos de terraços presentes na área de estudo correspondem a antigos terraços do Rio Tejo, localizados na margem direita do mesmo. Estes depósitos são constituídos essencialmente por areias, seixos e cascalheiras com intercalações de camadas de argilas siltosas e de argilas arenosas.

Terciário

Miocénico

- **M⁴_{VIIa} – “Areolas de Braço da Prata”** – Esta formação está representada a noroeste dos terraços fluviais e corresponde a uma alternância de grés finos, de areolas, de areias e de pequenos leitos de calcários margosos ou gresosos, muito fossilíferos.

Hidrogeologicamente a área de estudo localiza-se nas proximidades da Unidade Hidrogeológica da Bacia do Tejo-Sado, nomeadamente nas sub-unidades Aluviões do Tejo e Margem Direita (Almeida *et al*, 2000).

O sistema aquífero das Aluviões do Tejo é formado por materiais de origem fluvial, aluviões modernas (Holocénico) e terraços (Plistocénico). No vale do Tejo os depósitos aluvionares são de um modo geral, formados por areias e argilas, interestratificadas, com um depósito basal formado por areias, seixos e calhaus, os quais podem atingir 40m de espessura. Os terraços são constituídos por depósitos basais com seixos e calhaus, seguidos por um complexo interglaciário formado por areias e argilas.

O rio Tejo é o grande eixo longitudinal de drenagem deste sistema aquífero. O escoamento subterrâneo é feito em direcção ao Rio Tejo e ao longo da faixa das aluviões modernas, até ao estuário.

O sistema aquífero da Margem Direita é formado essencialmente por formações detríticas, nomeadamente arcoses, depósitos conglomeráticos, arenitos, argilitos e calcários margosos. Este sistema aquífero é constituído por um aquífero carbonatado e por um sistema aquífero mais profundo instalado em rochas detríticas, estes são predominantemente confinados a semiconfinados.

Importa referir que as formações aluvionares e os depósitos de terraço aflorantes na área de estudo apresentam elevada permeabilidade e elevada vulnerabilidade à poluição, uma vez que estamos na presença de depósitos aluvionares, constituídos essencialmente por uma alternância de areias e argilas, o que torna os níveis arenosos muito permeáveis.

2. METODOLOGIA E FUNDAMENTOS TEÓRICOS

2.1. METODOLOGIA

Este trabalho tem como objectivo propor uma alteração da Simulação Sequencial Directa (SSD) e testar uma metodologia para caracterizar espacialmente os teores em TPH (contaminante) na área de uma antiga refinaria entretanto desactivada.

A informação de partida consiste em sondagens verticais efectuadas desde a superfície até ao substrato rochoso. As sondagens foram amostradas por troços sempre que se observou nos trabalhos de campo alteração significativa das características dos solos (cor, textura, cheiro, etc.). Por essa razão os troços analisados são de comprimento irregular.

Cada amostra de solo foi analisada do ponto de vista sensorial nos 3 parâmetros: (1) cheiro; (2) cor; e (3) classe de reacção ao óleo. Este último é um procedimento patenteado pela Arcalis® e consiste em observar a extensão e cor da película colorida de hidrocarbonetos que aparece quando um solo contaminado é colocado numa tina de água. Estas variáveis sensoriais são representadas por uma variável categórica (informação *soft*) com as modalidades (0, 1, 2, 3, 4 e 5), sendo o 0 correspondente à ausência de evidências de contaminação e o 5 a máxima evidência de contaminação. Estas três variáveis *soft* são abundantes, porque foram feitas em todos os troços das sondagens, mas não traduzem um quantitativo de contaminação, apenas indicam de forma subjectiva uma hipotética contaminação ou ausência de contaminação por hidrocarbonetos não voláteis.

Uma fracção do total de amostras foi enviada para laboratório para determinações analíticas aos vários componentes dos quais se pode estimar o TPH pela soma das concentrações em hidrocarbonetos. A concentração em TPH é a informação *hard*, e como é uma determinação analítica tem apenas o erro associado ao ensaio laboratorial. Esta informação é menos expedita e mais cara, por isso é menos abundante.

A ideia da metodologia proposta é combinar estes dois tipos de informação numa metodologia integrada e consistente, sabendo de antemão que se existir uma informação secundária mais abundante do que a informação principal e se esta informação estiver correlacionada então existem vantagens em utilizar os dois tipos de informação.

A modelação geoestatística, combinando informação principal e secundária já é muito comum (Goovaerts, 1997; Soares, 2000). Quando a informação secundária é conhecida em toda a área de estudo (com a mesma resolução espacial do modelo que se pretende como final) e esta é do tipo contínuo, mesmo que tenha uma magnitude completamente diferente da variável principal, podem ser utilizados algoritmos baseados na krigagem com informação secundária, nomeadamente a krigagem com deriva externa, krigagem com médias locais ou a cokrigagem colocalizada. Se a variável secundária é mais abundante mas apenas é conhecida parcialmente, então pode-se fazer uma estimação por krigagem desta variável secundária e

depois utilizar o resultado para condicionar a variável principal. Esta última solução pode ser aplicada para variáveis secundárias de tipo contínuo ou categórico. Quando a variável secundária é de tipo categórico a simulação por campos de probabilidade (*Probability Field Simulation* – PFS) é uma alternativa comum. Um estudo anterior apresentado por Pereira *et al*, 2001, propôs para este conjunto de dados uma metodologia baseada na estimação por krigagem da variável categórica da informação secundária (agrupada em duas modalidades principais em vez das seis originais) e na simulação dos teores por simulação sequencial gaussiana (SSG). As imagens geradas por SSG foram simuladas de forma independente para cada modalidade da variável categórica, ou seja, forma simulados pares de imagens, umas relativas a valores baixos de TPH e outra relativa aos valores altos. Estes pares de imagens foram posteriormente cruzados com a imagem binária da variável categoria dando origem a uma imagem final de teores de TPH.

Este procedimento garante imagens simuladas de TPH condicionais aos valores das amostras, e cujos histogramas são condicionais aos dados de partida, assim como os variogramas. Todavia, como os pares de imagens de valores altos e baixos são simulados independentemente um do outro, após a junção das imagens observam-se transições bruscas de teores na interface entre a classe de valores altos e baixos estimadas o que não é realista.

Para obviar este inconveniente, neste trabalho propõe-se uma metodologia em duas etapas semelhante à que foi utilizada no primeiro estudo, primeiro a modelação das classes de análise sensorial por estimação (Soares, 1992) e depois a modelação dos teores em TPH por simulação (Goovaerts, 1997, Soares, 1992; 2000). Na modelação dos teores em TPH considerou-se uma nova variante do algoritmo SSD, proposto inicialmente por Soares, para que a geração dos valores simulados seja feita por amostragem de um histograma local em vez do procedimento proposto na formulação original do algoritmo que é a amostragem de um histograma global. Em cada iteração da SSD, correspondente à simulação de um nó (bloco), o histograma local é o histograma condicional de TPH à classe de análise sensorial prevalente no nó a simular. No algoritmo da SSD, a estimação da média e variância em cada localização a simular é feita por krigagem simples com médias locais onde as médias locais são os teores médios em TPH da classe da variável secundária obtida por estimação.

Na figura 2.1 apresenta-se o diagrama com a metodologia utilizada. O software utilizado em todas as fases foi o geoMS (Alexandre e Almeida, 1998) (modelação dos variogramas e estimação por krigagem), e uma versão modificada da simulação sequencial directa (SSD) (Soares, 2001; Ruben & Almeida, 2010) para a simulação a partir da reamostragem dos histogramas locais. A modelação é feita a 3D até à profundidade de 3,5m, em 7 horizontes de 0,5 metros.

A primeira etapa do estudo consistiu na análise bivariada entre as classes das análises sensoriais (cheiro, cor e classe de reacção ao óleo) e os teores em TPH, nos dados onde estes dois tipos de informação coexistem simultaneamente. Esta análise bivariada foi feita por histogramas condicionais e permitiu seleccionar a variável sensorial que exhibe melhor

correlação com os teores e redefinir (reduzir) o número de classes. Simultaneamente são registados os valores médios de TPH condicionais às classes da variável secundária seleccionada por horizonte.

As etapas seguintes são a modelação 3D da variável sensorial seleccionada (dados *soft*). Calculam-se e modelam-se variogramas experimentais multifásicos e estimam-se em profundidade estas classes por krigagem multifásica. Os valores de probabilidade obtidos são classificados numa variável categórica com as modalidades de partida. Este modelo de classes é validado por observação e sobreposição dos dados das sondagens, assim como pelas proporções obtidas.

Seguidamente é feita a modelação 3D dos teores em TPH (dados *hard* ou seja, informação proveniente de análises laboratoriais). Em primeiro lugar, foram calculados os variogramas experimentais e foi ajustado um modelo teórico de variograma. Estes variogramas são calculados e ajustados para todo o conjunto de dados de TPH e não por classes da variável secundária.

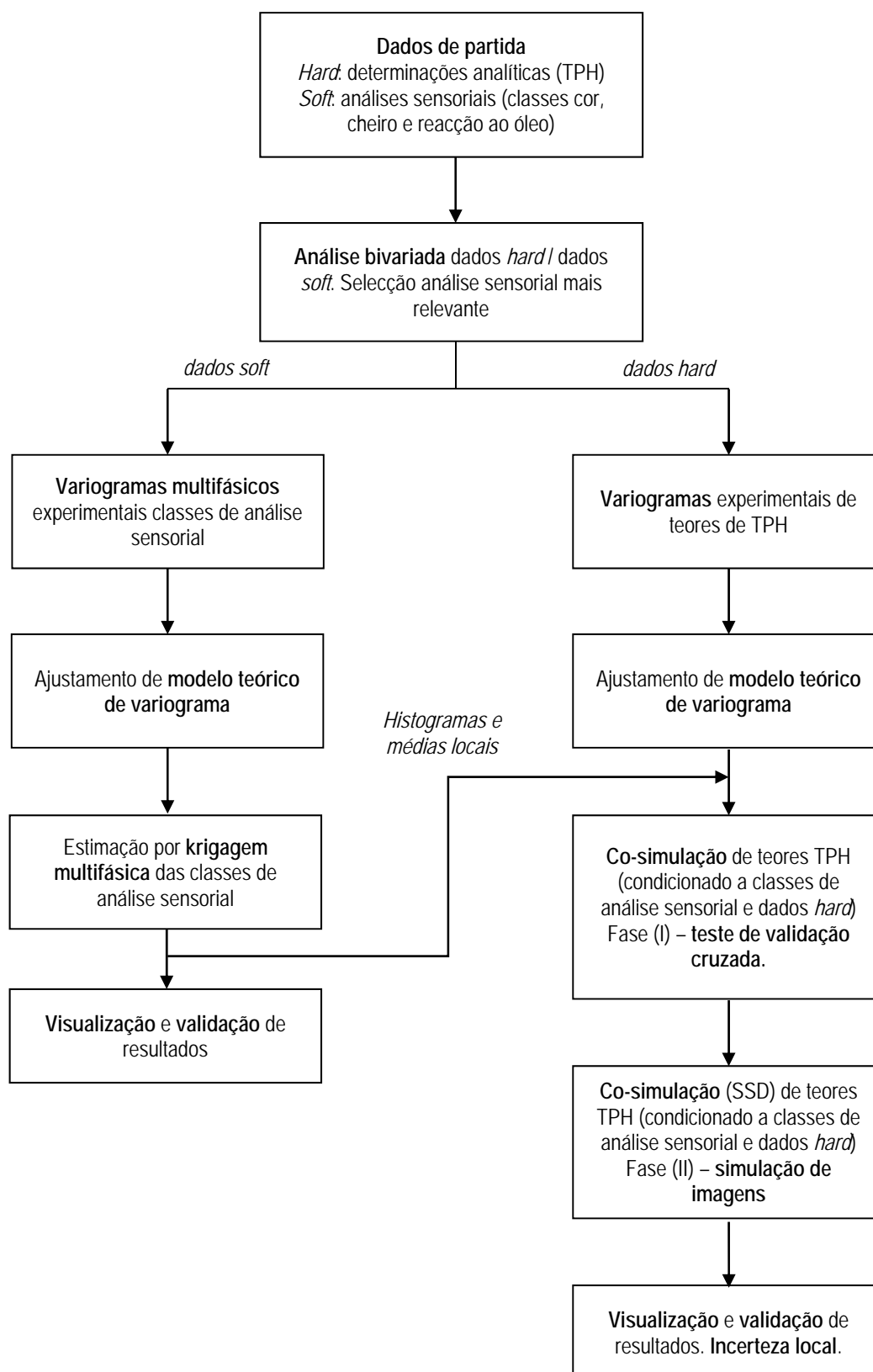


Figura 2.1 – Diagrama da metodologia proposta para a caracterização da contaminação em TPH na área da antiga refinaria em estudo.

Depois de modelados os variogramas experimentais, procedeu-se a um teste de validação cruzada. Neste teste, do conjunto de dados experimentais foram retiradas aleatoriamente 20% das amostras. Para cada variante do algoritmo de simulação, foram então geradas 100 imagens para todo o volume estudado. Seguidamente, retiveram-se os valores simulados para cada localização dos 20% de amostras previamente seleccionadas. Para cada conjunto de 100 valores simulado no local de cada amostra (do conjunto de 20%) é calculada a média (valor estimado). Para cada amostra é assim calculado um erro de estimação e um erro quadrático. Também, relativamente ao valor de corte 1000 ppm em TPH, foram determinadas o número de vezes em que a média dos valores simulados está acima e abaixo deste valor de corte. Foram testadas as seguintes variantes de simulação a que correspondem os acrónimos:

SSD_MG – simulação sequencial directa com médias globais (sem influência das classes de reacção ao óleo)

SSD_ML_CRO – simulação sequencial directa com médias locais por classe de reacção ao óleo e condicionamento a histograma global

SSD_HL_ML_CRO – simulação sequencial directa com médias locais por classe de reacção ao óleo e com condicionamento a histograma local

SSD_HL_ML_CRO&Z – simulação sequencial directa com médias locais por classe de reacção ao óleo e por horizonte, com condicionamento a histograma local.

Seguidamente, são simuladas 100 imagens de TPH com a variante escolhida. As imagens simuladas são validadas do ponto de vista do condicionamento aos dados experimentais, dos estatísticos e variogramas. Para terminar faz-se uma análise crítica da incerteza local que é mostrada pelo conjunto das amostras e apresentam-se os locais potencialmente contaminados para um hipotético valor de corte (*threshold*) de 1000 ppm.

2.2. FUNDAMENTOS TEÓRICOS GEOESTATÍSTICOS

A aplicação da geoestatística à área do ambiente, nomeadamente no âmbito da contaminação dos solos por hidrocarbonetos, é particularmente eficaz na medida em que se pretende caracterizar a dispersão espacial dos eventuais contaminantes. Dada a impossibilidade de avaliar de forma detalhada no espaço (lateralmente e em profundidade) a distribuição espacial de um contaminante, normalmente opta-se por efectuar a sua caracterização com um conjunto limitado de amostras, sendo a sua densidade determinada pelo comportamento mais ou menos errático das variáveis em estudo. Acresce ainda a combinação ou mistura entre medições *hard* ou primárias e medições *soft* ou secundárias, de forma a otimizar os custos de um plano de amostragem sem comprometer os resultados.

Um conjunto de dados amostrais tem sempre associada uma incerteza intrínseca, como sejam os erros resultantes da recolha propriamente dita, que neste caso de estudo são variáveis de tipo sensorial e analítica. A recolha dos dados apresenta características específicas, que ficam associadas a um grau de incerteza ligado ao conhecimento do fenómeno em estudo, consistindo no que se designa por uma realização espacial impossibilitando a repetição da observação no mesmo espaço e tempo.

A geoestatística é um ramo da estatística que lida com os problemas associados ao espaço (e/ou tempo), e tem como objectivo caracterizar a distribuição espacial (e/ou temporal) das grandezas medidas e apresentar a respectiva incerteza local e global, tendo em conta a densidade, heterogeneidade e zonamento da informação disponível (Soares, 2000).

Os modelos da geoestatística resultam na associação do comportamento da variável em estudo com os fundamentos teóricos da matemática e da estatística, em particular a Teoria das Funções Aleatórias, resultando no que se designa por um modelo probabilista.

Na geoestatística consideram-se duas estratégias de avaliação: estimação e simulação. Os modelos de estimação ou interpolação geoestatística são baseados no formalismo de krigagem e têm como objectivo obter uma imagem média do fenómeno (variável) em estudo. Existem várias variantes do estimador de krigagem conforme consideram ou não informação secundária (krigagem normal, co-krigagem, co-krigagem colocalizada, krigagem com deriva externa) ou sejam aplicados a variáveis do tipo categórico (krigagem da indicatriz e krigagem multifásica) ou contínuo. Os modelos de simulação têm como objectivo obter um conjunto de imagens equiprováveis que no seu todo quantificam a incerteza local e global. Qualquer uma destas abordagens necessita de ser validada à *posteriori*, mediante o cruzamento dos resultados obtidos com os dados das amostras e o conhecimento pericial e científico do sistema que está a ser estudado.

2.2.1. VARIÁVEIS ALEATÓRIAS

Uma variável aleatória é uma grandeza que tem associada um conjunto de valores distribuídos no espaço e/ou tempo, e que foi amostrada com uma determinada estratégia (Soares, 2000). Se forem designadas genericamente por x_i as localizações das amostras e por z_i as respectivas realizações, então $z(x_i)$ é a realização da variável aleatória na localização x_i .

Os fenómenos regionalizados assentam no conceito de variáveis regionalizadas e são interpretados como realizações das funções aleatórias que variam irregularmente no espaço. As variáveis regionalizadas são utilizadas na análise de fenómenos com variação espacial e/ou temporal e apresentam dois aspectos ambíguos sobrepostos: por um lado uma variação imprevisível de ponto para ponto do espaço, por outro, variações estruturadas, evidenciando correlação espacial.

As variáveis aleatórias são caracterizadas pelos dois primeiros momentos (média e variância). Entre duas variáveis aleatórias podem ser avaliadas estatísticas de correlação como a covariância e o coeficiente de correlação.

O conjunto de variáveis aleatórias $Z(x_i), i = 1, \dots, N$ correlacionadas entre si constituem uma função aleatória, do qual só se conhece uma realização $z(x_i)$, os dados amostrais. O facto de só se conhecer uma realização impossibilita o cálculo de parâmetros estatísticos das variáveis individuais ou da função.

Para ultrapassar este constrangimento, o formalismo geoestatístico assenta nas duas seguintes hipóteses:

- a) Estacionaridade da média ou de 1ª ordem – admite-se que todas as variáveis aleatórias têm a mesma média.

$$E\{Z(x_1)\} = E\{Z(x_2)\} = \dots E\{Z(x_i)\} = E\{Z(x)\} = m$$

Esta estatística passa a ser independente da localização x_i e pode ser estimada pela média aritmética dos valores das realizações das variáveis aleatórias:

$$m = \frac{1}{N} \sum_{i=1}^N Z(x_i)$$

- b) Estacionaridade da covariância espacial ou de 2ª ordem (continuidade espacial) – admite-se que a correlação entre duas variáveis aleatórias depende só da distância espacial que as separa (vector \vec{h}) e é independente da localização. Esta hipótese torna possível a estimação da covariância espacial (ou variograma) com base nas N amostras disponíveis.

$$C\{Z(x_1), Z(x_2)\} = C\{Z(x_i), Z(x_{i+h})\} = C(h)$$

ou

$$\gamma\{Z(x_1), Z(x_2)\} = \gamma\{Z(x_i), Z(x_{i+h})\} = \gamma(h)$$

A assunção destas duas hipóteses de estacionaridade da função (x) no modelo probabilista é o mesmo que dizer que se assume que o conjunto de amostras é homogéneo e representativo da área. Estas hipóteses não podem ser provadas ou refutadas pelos dados.

2.2.2. VARIOGRAFIA E AJUSTAMENTO DE MODELOS TEÓRICOS

Para se compreender a distribuição espacial característica de uma variável é necessário quantificar a sua continuidade espacial. Um dos elementos mais utilizados para a avaliação da continuidade espacial é o biponto, par de pontos separados por um vector \vec{h} , que permite avaliar o modo como a continuidade de X varia nas diferentes direcções do espaço (Soares, 2000). O elemento biponto é considerado um elemento estruturante de vasta aplicação da prática geoestatística de análise de continuidade espacial de um recurso natural, uma vez que

o conhecimento que se possui do recurso advém de um conjunto de amostras ou observações, regularmente discreto, limitado, e na maioria das vezes, escasso.

Para uma dada característica quantitativa Z , os diagramas de representação de pares de pontos (x) e $(x + h)$, calculados com base no biponto para diferentes valores de \vec{h} , são considerados como sendo os que contêm informação mais relevante sobre a continuidade espacial de $Z(x)$ (Soares, 2000). Tal diagrama de representação é denominado por *h-scattergram* (Goovaerts, 1997). No entanto, a análise da continuidade espacial com recurso aos *h-scattergrams* exige a realização de n diagramas, tantos quanto o número de $\vec{h}(s)$ que se pretende avaliar. Por este motivo, e para uma melhor interpretação e posterior utilização é necessário sintetizar a informação dos *h-scattergrams* em instrumentos que permitem visualizar a evolução dos dados com \vec{h} , quantificando assim, a continuidade de $Z(x)$.

A sintetização habitual dos *h-scattergrams* é feita com o chamado variograma $\gamma(h)$. Cada valor de um variograma resume a informação de um *h-scattergram* através da média do quadrado das diferenças entre todos os pares de pontos separados por \vec{h} . No gráfico *h-scattergram*, o valor do variograma corresponde à distância média de todos os pontos à recta a 45°.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

Nesta expressão, (h) é o número de pares de pontos para cada valor de h . O variograma experimental pode ser calculado por direcções, resultando numa análise multi-direccional que permite a identificação e quantificação de anisotropias.

A análise dos variogramas é condicionada pelo conhecimento prévio que se possui do fenómeno espacial em estudo. Da análise resultam as seguintes medidas (Pereira, 1979; Soares, 2000):

a – amplitude – distância a partir da qual as amostras se tornam independentes, ou seja, deixam de ter correlação entre si. Corresponde à noção intuitiva de zona de influência de uma amostra.

C – patamar – variância *a priori* da variável em estudo e reflecte a dispersão própria do fenómeno mineralizado para distâncias superiores à amplitude.

C_0 – efeito pepita – ordenada na origem do variograma e resulta da sobreposição de microestruturas e de erros de amostragem. Quantifica a variabilidade a uma escala inferior à da amostragem.

Fenómenos geológicos particulares podem apresentar continuidade variável consoante a direcção (Soares, 2000). Veja-se o exemplo de uma lentícula aurífera. É de esperar que apresente maior continuidade na direcção paralela ao seu eixo de maior expressão, do que na direcção perpendicular a esse mesmo eixo. Quando assim acontece, a estrutura/característica

diz-se anisótropa. Quando tal não se verifica, ou seja, quando o variograma apresenta o mesmo comportamento em todas as direcções do espaço, a mesma denomina-se de isotrópica.

Os valores dos variogramas para diferentes passos h , de uma dada variável (x), são representados graficamente em função do passo h . Estes valores são ajustados por uma função geral e representativa (ou soma de funções), com um número reduzido de parâmetros (efeito de pepita, patamar e amplitude). Com o ajustamento, infere-se um andamento de $\gamma(h)$ representativo para toda a área e para toda a gama de distâncias h . Esta acção é de vital importância, uma vez que permite a realização da síntese das características estruturais do fenómeno espacial, como sejam o grau de dispersão/continuidade, anisotropias, imbricamentos, num modelo único e coerente de variograma (Soares, 2000).

A escolha das funções a utilizar no ajustamento é limitada a um conjunto de modelos de funções definidas positivas, dos quais as mais utilizadas são o modelo esférico e o exponencial (respectivamente 2.1 e 2.2).

$$\gamma(h) = \begin{cases} C \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a} \right)^3 \right] & \text{para } h \leq a \\ C & \text{para } h > a \end{cases} \quad 2.1$$

$$\gamma(h) = C[1 - e^{-3h/a}] \quad 2.2$$

Refira-se que o modelo esférico atinge o patamar à distância a (amplitude); já o modelo exponencial nunca atinge o patamar, isto é o patamar é uma assintota. Neste último, para uma distância igual à amplitude, a função atinge aproximadamente 95% do patamar.

Alguns variogramas experimentais mostram o que se designa por estrutura imbricada, ou seja, mais de um patamar e amplitude, o que acontece quando coexistem diferentes escalas de variabilidade nos dados. Em tais situações para a representação do variograma torna-se necessário realizar combinações aditivas de modelos simples, denominados por modelos imbricados.

2.2.3. ESTIMAÇÃO POR KRIGAGEM

A krigagem é o nome geral dado ao conjunto de processos de estimação. Através da utilização de estimadores de krigagem pode-se estimar valores de uma variável em localizações não amostradas a partir de observações assumidamente pontuais. Este formalismo foi apresentado pela primeira vez pelo francês G. Matheron (1963) com base nos trabalhos desenvolvidos na área da geologia pelo Sul-africano Daniel G. Krige que foi o pioneiro a introduzir o uso de médias móveis para evitar a sobrestimação sistemática de reservas de ouro.

O procedimento de krigagem baseia-se numa interpolação por média móvel ponderada, contudo, ao contrário de outros métodos de interpolação por distância, os pesos atribuídos às

amostras (λ_i), são determinados a partir de uma análise espacial baseada no variograma experimental, de forma a minimizar a variância estimada. O uso do variograma experimental na atribuição dos pesos garante de que as características estruturais e de homogeneidade do processo a estimar são levadas em consideração uma vez que, como já referido, o variograma mede a continuidade e a homogeneidade de um processo espacial do qual conhecemos apenas um conjunto de amostras.

O estimador de krigagem é um estimador BLUE (*Best Linear Unbiased Estimator*) pois visa a obtenção do melhor estimador linear possível, a partir da combinação linear dos valores observados do fenómeno, isto é: *linear*, porque as suas estimativas são combinações lineares ponderadas dos dados existentes; *não enviesado*, pois procura que a média dos erros de estimação (desvios entre o valor real e o valor estimado) seja nula; e *o melhor*, porque os erros de estimação apresentam uma variância (variância de estimação) mínima.

Para diferenciar alguns métodos de krigagem, Goovaerts (1997), admite que a função aleatória (x) se decompõe em duas componentes aditivas, uma componente designada por tendência ou deriva (x) e uma componente residual (x).

$$Z(x) = M(x) + R(x)$$

Assumindo que a componente residual tem média nula, o valor da função (x) será igual à componente tendência (x). Segundo o mesmo autor, o modelo adoptado para a componente tendência permite distinguir os três estimadores de krigagem linear mais usuais: krigagem simples, normal ou ordinária e a universal.

A krigagem simples assume a hipótese de estacionaridade da média, ou seja, assume a mesma média m para toda a área de estudo. Esta hipótese permite fazer a estimação de uma localização desconhecida através dos chamados resíduos:

$$Y(x_i) = Z(x_i) - m$$

O estimador de $Z(x)$ na localização x_0 à custa de P amostras vizinhas é dado por:

$$[z(x_0)]^* = m + \sum_{i=1}^P \lambda_i \cdot [z(x_i) - m]$$

A krigagem normal não assume a hipótese de estacionaridade da média, ou seja, assume a média m na localização a estimar como desconhecida. O estimador de $Z(x)$ na localização x_0 é dado por:

$$[z(x_0)]^* = \sum_{i=1}^P \lambda_i \cdot z(x_i)$$

Para qualquer formulação de krigagem, os ponderadores λ_i são obtidos pela resolução de um sistema de krigagem formulado à custa das covariâncias para as distâncias entre as amostras e entre as amostras e a localização a estimar (Soares, 2000). No caso da krigagem normal assume-se ainda que a soma dos ponderadores é igual a um (condição de não enviesamento).

2.2.4. FORMALISMO DA INDICATRIZ

Na metodologia proposta faz-se em primeiro lugar a estimação das classes de análise sensorial que é uma variável categórica. A modelação geoestatística de uma variável categórica recorre ao chamado formalismo da indicatriz simples (2 modalidades) (Journel e Isaaks, 1984) ou multifásico (mais de duas modalidades) (Soares, 2000; Almeida; 2010).

O formalismo da indicatriz baseia-se na transformação dos valores categóricos para uma variável binária do tipo zero e um. Por exemplo, se assumirmos duas modalidades, solo contaminado e solo não contaminado, uma variável $z(x)$ pode ser transformada numa variável indicatriz $I(x)$ do seguinte modo:

$$I(x_i) = \begin{cases} 0 & \text{se } z(x) < z - \text{Solo não contaminado} \\ 1 & \text{se } z(x) \geq z - \text{Solo contaminado} \end{cases} \quad 2.3$$

Através deste classificador, é possível transformar os dados originais numa população binária pertencente à classe X e ao seu complementar X^c numa área A , em que o conjunto é representado por $A = X \cup X^c$.

2.2.4.1. POPULAÇÕES MULTIFÁSICAS COM P CLASSES

As populações multifásicas são constituídas por diferentes classes X e X^c , como por exemplo determinada litologia, um tipo de minério, uma determinada espécie, ou um intervalo de valores de uma variável qualitativa, como é o caso do presente estudo, mais de duas classes de análise sensorial. Quando assim acontece, ou seja, quando a população pode ser decomposta em p classes (fases segundo Soares, 2000 e Luís, 2004), tal que $I(x_i) = \text{prob}\{x_i \in X_k\}$ com $k = 1, \dots, p$ e $A = \{X_1 \cup X_2 \cup \dots \cup X_p\}$, qualquer ponto x localizado em A passa a poder ser definido por um vector binário $I_k(x)$ tal que:

$$I_k(x_i) = \begin{cases} 1 & \text{se } x_i \in X_k \\ 0 & \text{se } x_i \notin X_k \end{cases}, \quad X_k (k = 1, \dots, p) \quad 2.4$$

Em que X_k representa cada classe definida em A .

Segundo Luís, 2004, a variável indicatriz expressa a probabilidade de pertença de cada ponto x_i a cada fase X_k , podendo ser considerada como uma realização de uma função aleatória $I(x)$ cuja lei de distribuição binomial pode ser caracterizada, em cada fase, pelos seguintes momentos de primeira e segunda ordem:

$$m_k = E\{I_k(x)\} \quad 2.5$$

$$\sigma_k^2 = \text{var}\{I_k(x)\}, k = 1, \dots, p \quad 2.6$$

Quando se pretende estimar a probabilidade de pertença de uma variável a qualquer fase X_k , é possível a utilização de uma combinação linear do conjunto de valores da indicatriz das amostras experimentais vizinhas, permitindo assim a aplicação dos processos de estimação geoestatísticos, segundo o formalismo da indicatriz (Luís, 2004). Deste modo, a transformação dos dados, permite minimizar o normal desvio sistemático das funções de distribuição de probabilidades.

2.2.4.2. VARIOGRAFIA DE POPULAÇÕES MULTIFÁSICAS COM P CLASSES

Na análise da continuidade espacial para populações multifásicas, ou seja, com p classes, recorre-se à covariância espacial e/ou variogramas. A covariância de uma população multifásica pode ser interpretada como a probabilidade de duas amostras localizadas em $(x_i$ e $x_i + h)$, pertencerem à mesma fase X_k , qualquer que esta seja, $k = 1, \dots, K$, e a mesma pode ser decomposta na soma das covariâncias individuais (Almeida, 1999, Soares, 2000):

$$C(h) = \sum_{k=1}^K E\{I_k(x) \cdot I_k(x+h)\} = \sum_{k=1}^K C_k(h) \quad 2.7$$

O mesmo procedimento pode ser realizado para o cálculo do variograma de uma população multifásica.

$$\gamma(h) = \frac{1}{2} E \left\{ \sum_{k=1}^K [I_k(x) - I_k(x+h)]^2 \right\} = \sum_{k=1}^K \gamma_k(h) \quad 2.8$$

Na modelação dos variogramas, o ideal será a obtenção de um único modelo através da soma dos variogramas individuais (Soares, 2000). No entanto, dependente do número de amostras e classes definidas, a possibilidade de obtenção de um único modelo através da soma dos individuais fica condicionada, sendo a sua impossibilidade tanto mais acentuada quanto maior for o número de classes. Quando assim acontece, a caracterização deve ser realizada com recurso a mais do que um variograma multifásico, mas ainda assim, não mais do que dois (Almeida, 1999). Considerando (h) como o número de pares de pontos separados por h , o variograma multifásico correspondente pode ser obtido directamente pela seguinte expressão:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left\{ \sum_{k=1}^K [I_k(x_i) - I_k(x_i+h)]^2 \right\} \quad 2.9$$

A avaliação da continuidade espacial prossegue com a análise das anisotropias, caso existam, nas diferentes direcções, ajuste experimental de uma função teórica, e análise dos respectivos parâmetros de síntese.

Importa referir que os parâmetros que forem seleccionados vão condicionar todos os resultados futuros. A amplitude reflecte uma média da dimensão das diferentes classes ou corpos, sendo que as relações de anisotropias traduzem a variação dessas mesmas dimensões nas diferentes direcções principais. O efeito de pepita é uma medida reveladora da irregularidade morfológica no contacto entre os diferentes corpos (Soares, 2000), correspondente à heterogeneidade a pequena escala, ou seja, para curtas distâncias. Tal é reflectido nas imagens resultantes da simulação dos dados. Quando o efeito de pepita é elevado, a imagem resultante apresenta um número significativo de transições à pequena escala, e considerável irregularidade no contorno dos corpos (Almeida, 1999; 2010).

Em suma, tanto a covariância (h) como o variograma multifásico (h) quantificam a variabilidade morfológica média das estruturas multifásicas (Soares, 2000), que é um dos resultados a reter no presente trabalho.

2.2.5. ESTIMAÇÃO DA VARIÁVEL INDICATRIZ MULTIFÁSICA

A estimação da variável indicatriz multifásica, também chamada krigagem morfológica, é a extensão do sistema de krigagem para as variáveis do tipo indicatriz com K fases (modalidades) à custa de N amostras vizinhas

$$[I_k(x_0)]^* = \sum_{i=1}^N \lambda_i \cdot I_k(x_i) \quad 2.10$$

Os valores estimados $[I_k(x_0)]^*$ representam a probabilidade da localização x_0 pertencer à modalidade k . Após a estimação destes valores para todas as localizações e fases, é necessário transformar estes valores de probabilidade numa categoria. O método mais habitual é fazer a classificação baseada na optimização local dos valores de probabilidade, condicionados às estatísticas globais estimadas (Soares, 1992; Almeida *et al*, 1993).

2.2.6. SIMULAÇÃO SEQUENCIAL DIRECTA COM CONDICIONAMENTO A HISTOGRAMAS LOCAIS

Com a simulação geostatística pretende-se gerar imagens capazes de reproduzir as características do fenómeno em estudo, como sejam a lei de distribuição, a variabilidade e a continuidade espacial. Obtêm-se várias imagens que apresentam comportamentos estatísticos iguais ao dos dados das amostras, e uma relação exacta com a informação secundária/condicionante (Goovaerts, 1997; Soares, 2000; Luís, 2004). Estes resultados são complementares ao que se pretende com a estimação, que é a caracterização das distribuições espaciais das características médias de um recurso natural. Estas características tornam a simulação capaz de construir diferentes comportamentos, inclusive os extremados, para as características de dispersão e concentração da variável em estudo, à custa de um amplo conjunto de imagens.

Existem vários algoritmos para simulação, mas no presente trabalho seleccionou-se a Simulação Sequencial Directa (SSD). O princípio da simulação sequencial é bastante simples e baseia-se num condicionamento crescente à informação existente na vizinhança dos pontos a simular, que é obtido, tanto pelos valores experimentais, como pelos valores anteriormente simulados. Este procedimento é realizado em passos sequenciais sucessivos, e de acordo com a relação de Bayes:

$$F(Z_1, Z_2, Z_3, \dots, Z_N) = F(Z_1) F(Z_2|Z_1) F(Z_3|Z_1, Z_2) \dots F(Z_N|Z_1, Z_2, Z_3, \dots, Z_{N-1}) \quad 2.11$$

Considere-se a lei de distribuição conjunta de N variáveis aleatórias, e n valores condicionantes experimentais, $F(N) = (Z_1, Z_2, Z_3, \dots, Z_N|n)$. O processo de simulação genérico de (N) é passível de ser resumido nas seguintes etapas:

- i. Simulação de um valor z_1 , considerando a função de distribuição cumulativa $F = (Z_1|n)$, a partir de um caminho aleatório, escolhido sobre o campo geométrico em estudo. Uma vez realizada a simulação de z_1 , o valor obtido é adicionado ao conjunto de informação condicionante, para os subseqüentes passos de simulação, passando n a $n + z_1$;
- ii. Simulação de um novo valor z_2 a partir da lei condicional Z_2 , tendo por base os $(n + 1)$ valores condicionantes. Finda a simulação de z_2 , os dados condicionantes passam a ser actualizados para $(n + 2) = (n + 1) + z_2$;
- iii. Repetição do processo sequencial até se proceder à simulação completa das N variáveis.

Saliente-se que, as N variáveis aleatórias dependentes $Z_1, Z_2, Z_3, \dots, Z_N$, podem ser correspondentes à mesma característica, espacialmente referenciada nos N nós do campo espacial em estudo. Caso se considerem os n valores condicionantes iniciais como correspondentes aos valores experimentais, então $F(N) = (Z(x_1), Z(x_2), Z(x_3), \dots, Z(x_N)|(n))$.

No entanto, é no desconhecimento das N funções cumulativas condicionais, em casos práticos, que se centra a dificuldade de aplicação da simulação sequencial.

$$\begin{aligned} & \text{prob}\{Z(x_1) < z|(n)\} \\ & \text{prob}\{Z(x_2) < z|(n + 1)\} \\ & \text{prob}\{Z(x_3) < z|(n + 2)\} \\ & \vdots \\ & \text{prob}\{Z(x_N) < z|(n + N - 1)\} \end{aligned} \quad 2.12$$

É a necessidade do conhecimento destas funções que limita a aplicação prática destes algoritmos. No entanto, Journel e Alabert, 1989, ou Soares, 2001, consideram a utilização da krigagem como a ferramenta adequada na estimativa dessas, particularmente a krigagem multigaussiana para a simulação sequencial gaussiana (SSG), a krigagem de indicatriz para a simulação sequencial da indicatriz (SSI), e ainda a estimação por krigagem simples, com ou sem variáveis secundárias para a simulação sequencial directa (SSD) (Soares, 2001).

Se designarmos por (x) a variável a simular, a Simulação Sequencial Directa é resumida nas seguintes etapas (Soares, 2001):

- a) Escolha de uma sequência aleatória de nós de uma malha regular a simular, que permita uma única visita a cada nó.
- b) Em cada localização a simular x_u simulação do valor de $z^s(x_u)$, utilizando o algoritmo de Simulação Sequencial Directa:
 - Estimação da média e variância de (x) , $z(x_u)^*$ e $\sigma_{sk}^2(x_u)$ por krigagem simples.
 - Re-amostrar localmente o histograma de (x) , utilizando, por exemplo, uma transformação gaussiana (φ) da variável $Z(x)$; cálculo de $[y(x_u)]^* = \varphi[z(x_u)^*]$;
 - Geração de um número aleatório p , a partir de uma lei uniforme U entre $[0,1]$;
 - Geração do valor y^s a partir de $G(y(x_u)^*, \sigma_{sk}^2(x_u))$:

$$y^s = G^{-1}(y(x_u)^*, \sigma_{sk}^2(x_u), p);$$
 - Retorno do valor simulado da variável primária $z^s(x_u) = \varphi^{-1}(y^s)$.

A metodologia apresentada neste estudo no ponto 2.1 evidencia as seguintes alterações a este desenvolvimento geral da SSD:

- a) Na estimação da média e da variância locais de $Z(x)$, $z(x_u)^*$ e $\sigma_{sk}^2(x_u)$ por krigagem simples considera-se a informação secundária (classes de análise sensorial $k, k = 1, \dots, K$) estimada previamente para a malha em estudo. Ou seja, em cada localização a estimar x_u é conhecida a classe de análise sensorial mais provável $k(x_u)$. Esta malha de informação secundária (classes de análise sensorial) permite definir uma malha de médias locais na variável principal (neste estudo o TPH) correspondentes aos valores médios por classe de análise sensorial $m_{Z_k}(x_u), k = 1, \dots, K$:

$$m_{Z_k}(x_u) = \frac{\sum_{i=1}^{N_s} Z(x_i)}{N_s} \text{ se } K(x_i) \in k, k = 1, \dots, K$$

Com N_k igual ao número total de amostras com informação principal $Z(x)$ e secundária $K(x)$ simultaneamente e com a condição $K(x_i) \in k$.

- b) Para histogramas locais consideram-se não os histogramas globais da variável (x) , mas os histogramas condicionais de $Z(x)$ às modalidades da variável secundária.

2.3. TESTES DE VALIDAÇÃO CRUZADA

Um Teste de Validação Cruzada em geoestatística tem como objectivo aferir a qualidade de uma solução de modelo, incluindo variograma e estimação (Soares, 2000).

Consiste na eliminação sucessiva do valor de uma dada amostra ou conjunto de amostras do conjunto de amostragem com posterior estimação do seu valor, através de um algoritmo de krigagem ou simulação, à custa dos valores das amostras vizinhas. O conhecimento do valor

real das amostras retiradas permite a obtenção de um conjunto de pares – valores reais vs. valores estimados – e sobre esses pares efectuar o cálculo de alguns estatísticos de base como sejam: o erro médio EM (medida de enviesamento) e o erro quadrático médio EQM (medida da magnitude dos erros em valor absoluto). Idealmente tanto o EM como o EQM devem ser próximos de zero.

3. ANÁLISE ESTATÍSTICA PRELIMINAR

À área de estudo que foi utilizada para implementar e testar a metodologia de caracterização proposta no âmbito deste estudo corresponde a um antigo parque de combustíveis que se estende por uma área de aproximadamente 14,5 ha (figura 3.1). Como já foi referido anteriormente com mais detalhe, no local foram encontradas valas de drenagem, tanques, tubagens, áreas de trasfega, etc. que ao longo de décadas muito contribuíram para a contaminação evidenciada no local.

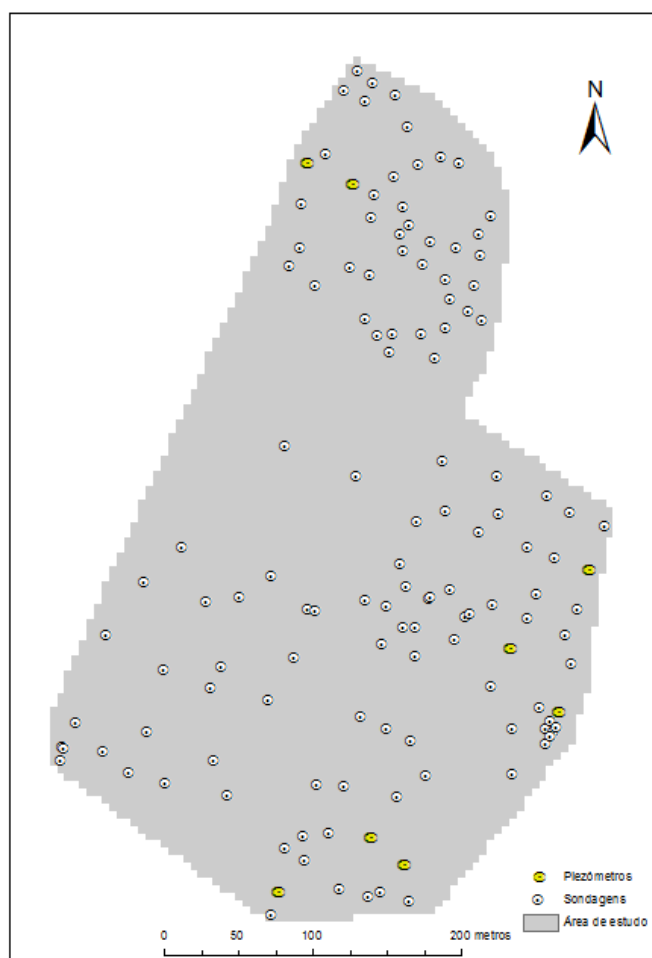


Figura 3.1 – Planta de localização das sondagens e piezómetros e discretização da área de estudo numa malha de blocos de 5 x 5 x 0,5 metros.

Foram utilizados dados de 131 furos realizados uma sonda com recuperação de testemunho (amostra não destrutiva) cuja força era aplicada por um martelo eléctrico de percussão. Este tipo de sonda permite recuperar amostras de camadas específicas e obter perfis detalhados

dos furos. Dos 131 furos, sete atingiram maior profundidade e foram transformados em piezómetros. As análises à água não foram todavia consideradas neste estudo.

As sondagens atingiram profundidades muito variadas, entre 0,80 e 8 metros, mas a maioria atinge 3 metros exactos. O histograma de profundidades máximas encontra-se representado na figura 3.2.

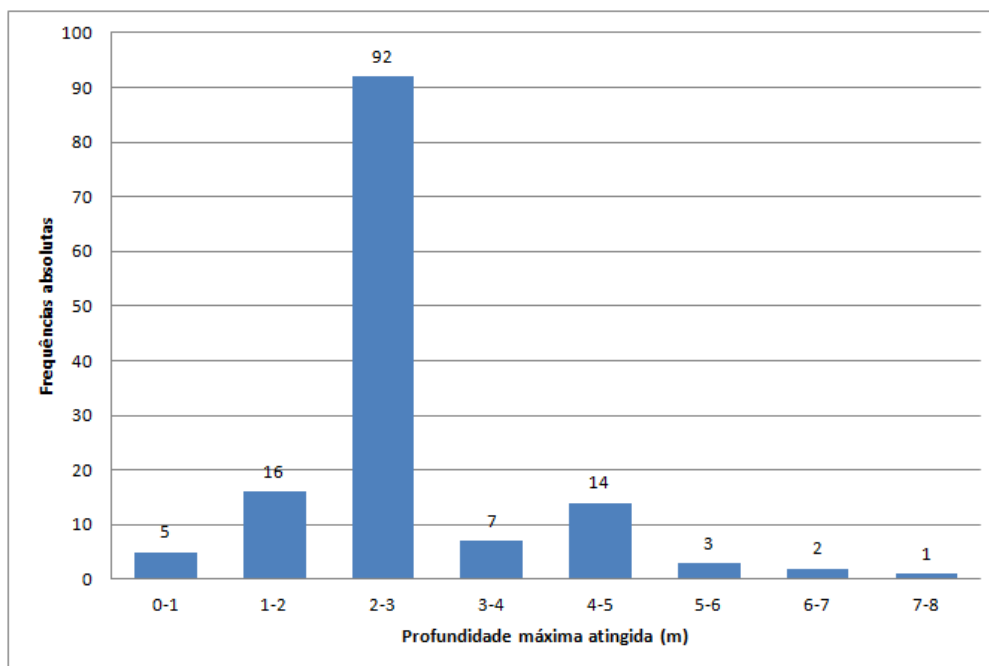


Figura 3.2 – Histograma de profundidade máxima atingida pelas sondagens.

Os solos retirados das sondagens foram classificados por classes de análise sensorial no que respeita à cor, cheiro e classe de reacção ao óleo. O comprimento dos suportes das amostras é muito variado. Para as classes de análise sensorial, os comprimentos da amostra variaram entre 5cm e 5 metros. Ao todo foram classificados 478 troços de sondagem perfazendo 42835 metros. Destes, amostras de 101 troços foram enviadas para laboratório para determinações analíticas (9740 metros, ou seja, 22,74%). A figura 3.3 mostra o histograma dos comprimentos dos troços que foram amostrados em cada caso.

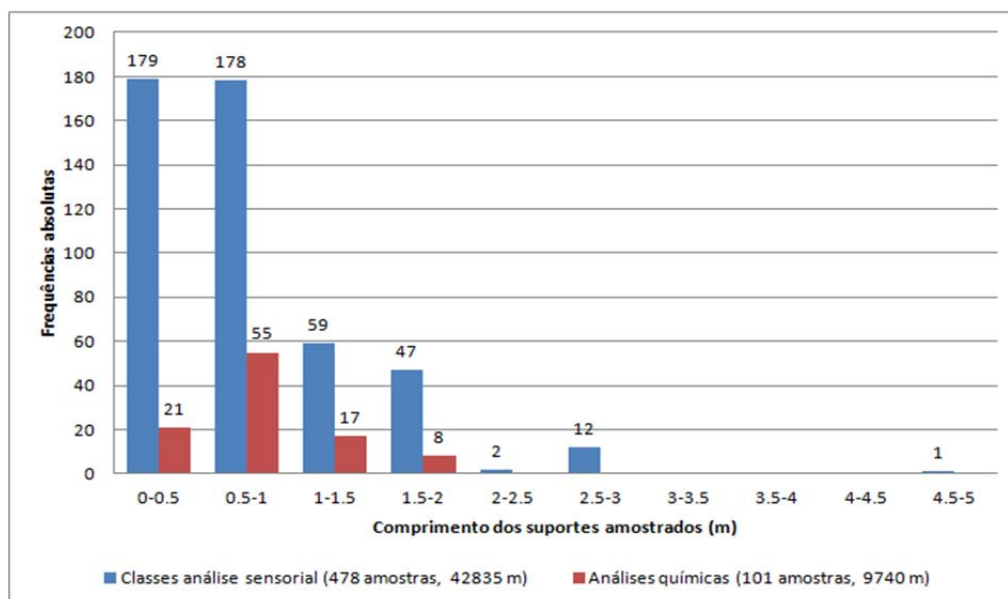


Figura 3.3 – Histograma dos comprimentos dos suportes amostrados em termos de classe sensorial e análises químicas.

A figura 3.4 mostra o número de amostras que foram classificadas em cada classe de análise sensorial. Essencialmente, observa-se que o número de troços de classe zero em todas as análises sensoriais é mais predominante, o que significa que o comprimento amostrado não contaminado prevalece. Outro aspecto a retirar do histograma, é que a análise sensorial cor dá mais resultados do tipo não contaminado (classe zero) e que a classe de reacção ao óleo aparece mais distribuída por todas as classes, mesmo assim predominam os valores na classe zero. Em termos genéricos, pode-se referir que a análise classe de reacção ao óleo é mais fidedigna na detecção de contaminação e que a cor é a característica sensorial mais optimista, sendo o cheiro a análise intermédia. Esta análise é todavia preliminar, e não tem em conta o cruzamento com os teores em TPH.

Para confirmar esta análise ao sentido optimista e menos fidedigna do resultado das análises sensoriais, apresenta-se a seguinte tabela bi-histograma, onde são apresentadas as frequências absolutas cruzadas das classes de reacção ao óleo e o cheiro e cor.

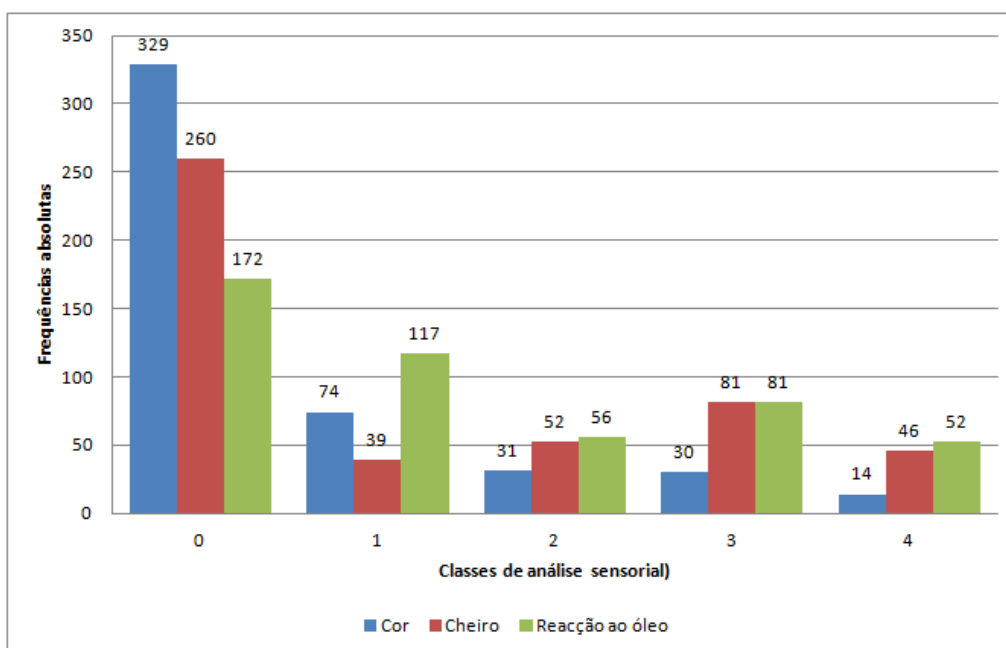


Figura 3.4 – Histograma do número de amostras por classe e tipo de análise sensorial.

Tabela 3.1 – Frequências absolutas cruzadas entre as classes de reacção ao óleo e as classes de cor e cheiro.

		Cor					Cheiro				
		0	1	2	3	4	0	1	2	3	4
Classes de reacção ao óleo	0	171	1	0	0	0	168	4	0	0	0
	1	110	7	0	0	0	86	29	2	0	0
	2	27	21	6	2	0	5	6	43	2	0
	3	12	37	19	13	0	1	0	6	71	3
	4	9	8	6	15	14	0	0	1	8	43

A tabela 3.1 mostra que 44% dos resultados das classes são coincidentes para a cor e reacção ao óleo, e que este valor sobe para 74% de coincidência quando a comparação é com o cheiro. O facto de predominarem frequências superiores a zero abaixo da diagonal, confirma a perspectiva menos fidedigna das análises sensoriais e cheiro relativamente à reacção ao óleo, principalmente a primeira.

Seguidamente procede-se à análise do grau de discriminação entre os teores de TPH e as classes de análise sensorial. Para esta análise construíram-se histogramas cumulativos condicionais dos teores em TPH (figuras 3.5, 3.6 e 3.7) por classes e calcularam-se estatísticos básicos condicionais (tabela 3.2). Todos os valores registados abaixo do nível de detecção (5 ppm) foram utilizados como tendo o valor 5 ppm.

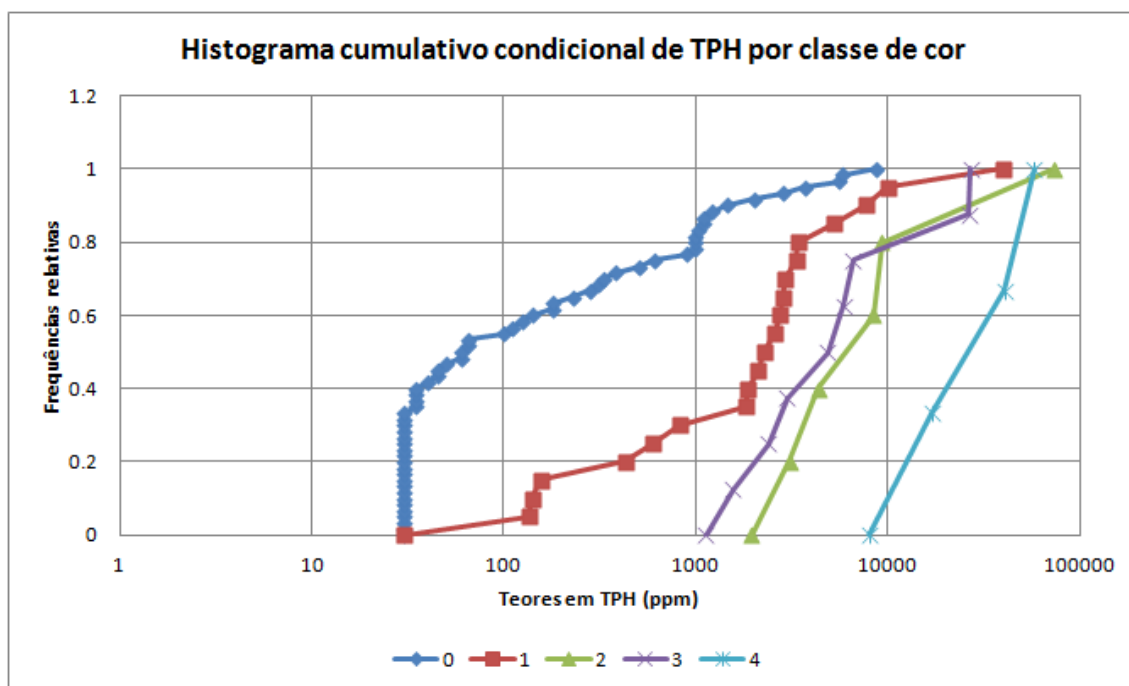


Figura 3.5 – Histograma cumulativo condicional dos teores em TPH por classe de cor.

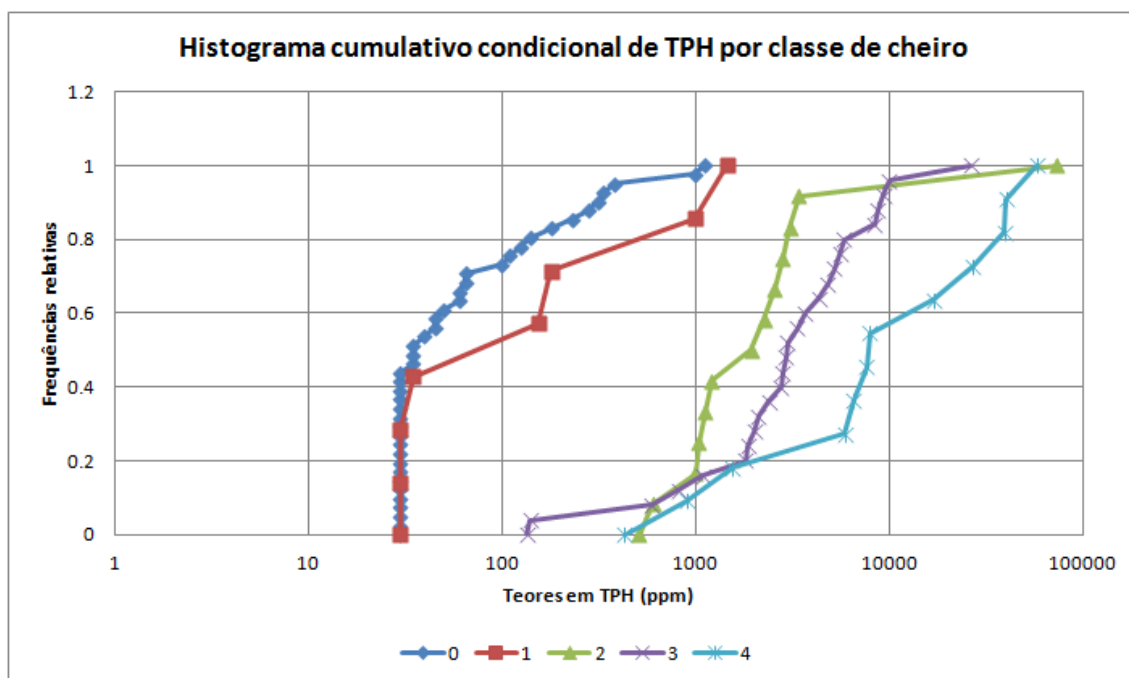


Figura 3.6 – Histograma cumulativo condicional dos teores em TPH por classe de cheiro.

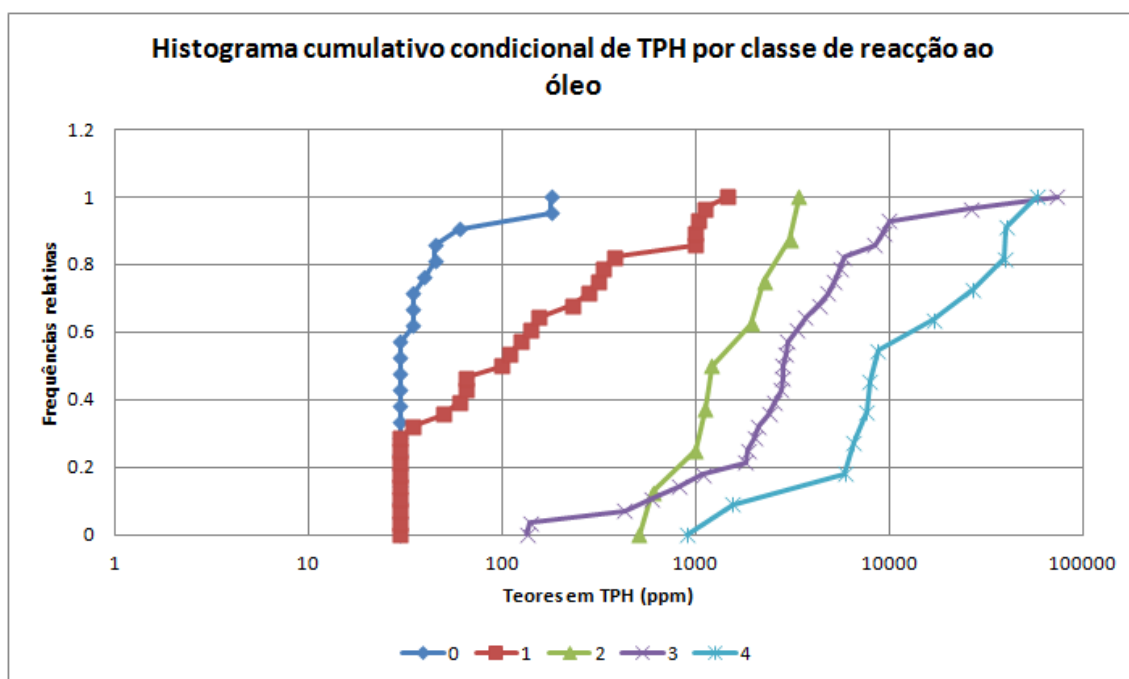


Figura 3.7 – Histograma cumulativo condicional dos teores em TPH por classe de reacção ao óleo.

Tabela 3.2 – Estatísticos básicos condicionais de TPH (ppm) por classe de análise sensorial.

Classes de análise sensorial	Classes	Nº de amostras	Média	Mediana	Desvio Padrão	Coef. de assimetria
Cor	0	61	709,59	60,00	1583,00	3,53
	1	21	4330,71	2275,00	8492,27	3,97
	2	6	16830,00	6382,50	28088,12	2,39
	3	9	8781,11	4850,00	10381,42	1,48
	4	4	30962,50	28772,50	22796,54	0,37
Cheiro	0	42	128,69	35,00	228,89	3,43
	1	8	360,63	90,00	549,25	1,61
	2	13	7343,46	1945,00	20006,19	3,59
	3	26	4637,31	2965,00	5279,73	3,06
	4	12	17811,67	7852,50	19160,06	1,08
Reacção ao óleo	0	22	47,50	30,00	43,53	2,91
	1	29	286,72	100,00	406,62	1,75
	2	9	1686,11	1210,00	1056,25	0,64
	3	29	6601,21	2840,00	13865,79	4,46
	4	12	18510,00	8407,50	18613,88	1,12
TODAS	01234	101	4337,52	595,00	11274,33	4,31

A análise dos gráficos e da tabela mostram que as 3 análises sensoriais discriminam os teores, e que regra geral os teores mais baixos encontram-se nas classes de menor grau e que os teores mais altos encontram-se nas classes de grau mais elevado.

Das 3 análises sensoriais observadas, a reacção ao óleo é a análise que subdivide os teores das classes mais baixas em subconjuntos com a menor dispersão (ver desvio padrão), sendo este facto mais evidente nas classes de índice 0, 1 e 2. Também as médias e medianas dos valores de TPH para cada classe são sempre crescentes, ao contrário das restantes análises sensoriais onde existem contradições. Por exemplo, a mediana que é resistente a valores anómalos, a classe 3 tem valor inferior à classe 2 o que representa uma contradição (má discriminação dos teores nestas classes). Os gráficos mostram também estas contradições, em vários valores.

Assim, em termos de discriminação dos teores, posicionou-se a reacção ao óleo como a melhor análise, seguido do cheiro e por fim da cor. Desta forma, seleccionou-se como variável secundária a reacção ao óleo.

Dado que as classes de modalidade mais elevada são pouco frequentadas, optou-se por agrupar as classes de reacção ao óleo 2, 3 e 4 numa única classe. Para as classes 0 e 1, optou-se por testar as classes em separado ou agrupadas. O histograma cumulativo condicional em teores de TPH (ppm) das novas classes (0, 1 e 2+3+4) e (0+1 e 2+3+4) encontra-se representado nas figuras 3.8 e 3.9. Procedeu-se à preparação dos ficheiros com estes novos agrupamentos.

Atendendo à distribuição em profundidade observada para os dados e para os teores em TPH, optou-se por construir um modelo de blocos, para georeferenciação do modelo geoestatístico, para uma profundidade máxima de 3,5m e uma discretização vertical de 0,5m. A tabela 3.3 mostra as características na malha de blocos nas 3 direcções ortogonais.

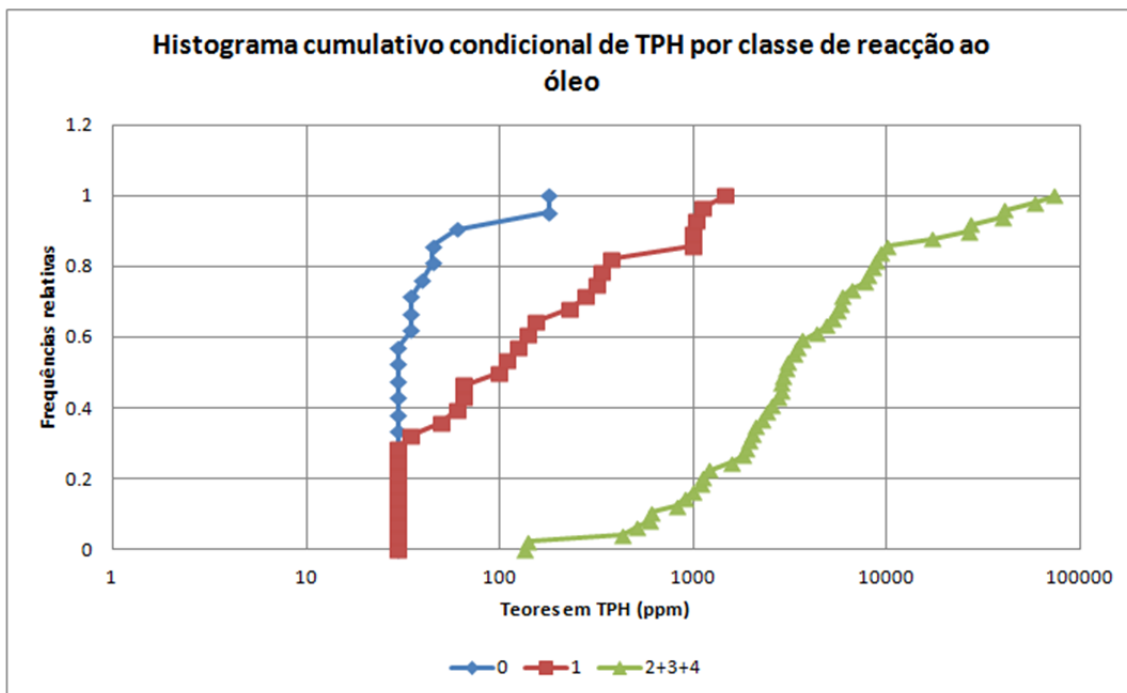


Figura 3.8 – Histograma cumulativo condicional dos teores em TPH por classe de reacção ao óleo após agregação das classes 2+3+4.

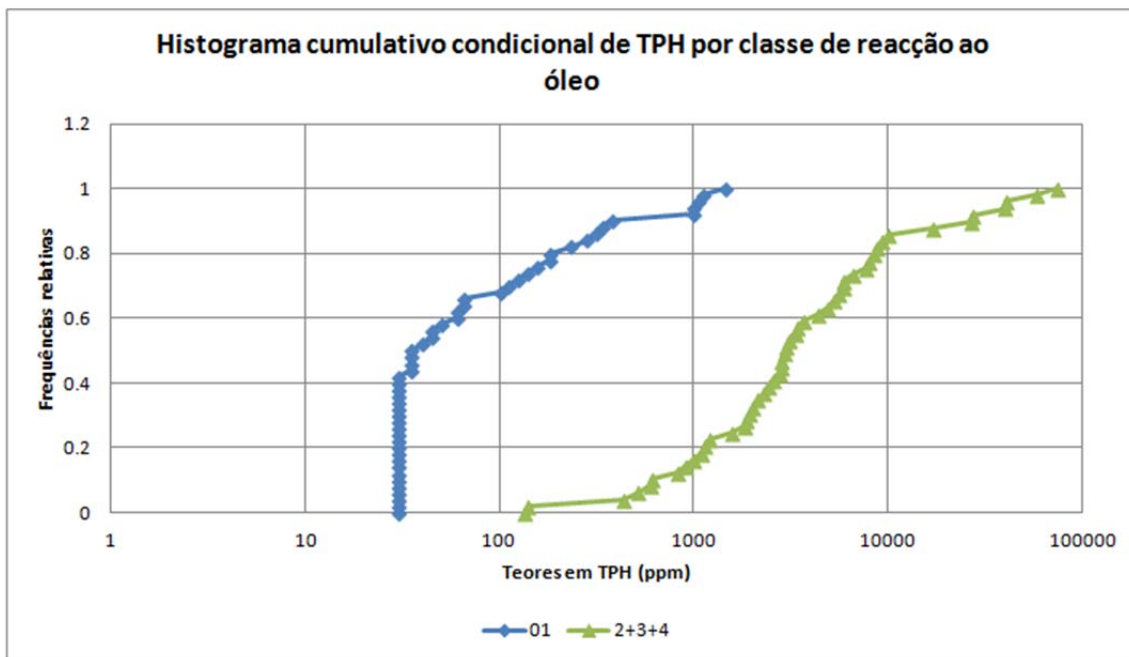


Figura 3.9 – Histograma cumulativo condicional dos teores em TPH por classe de reacção ao óleo após agregação das classes 0+1 e 2+3+4.

Tabela 3.3 – Posicionamento da malha de blocos para o modelo geoestatístico.

	Mínimo (m)	Número de blocos	Espaçamento (m)
X	-83750,00	77	5
Y	-95980,00	117	5
Z	0,50	7	0,5

Na preparação dos ficheiros de dados (classes e teores) teve-se em conta a discretização vertical de 0,5m. Os valores das classes foram calculados por mistura ponderada dos troços. Para os teores em TPH foi uma alocação dos valores por correspondência directa ao suporte, sem ponderação de teores. Por exemplo, um teor em TPH referenciado ao intervalo 0,70; 1,60, afecta com o mesmo valor os suportes [0,50; 1,00m] e [1,00; 1,50m]; os 10 cm de amostra não se consideram suficientes para caracterizar um suporte de 0,5m.

A título de exemplo mostram-se as primeiras amostras do resultado da mistura ponderada dos troços (tabela 3.4):

Tabela 3.4 – Extracto do ficheiro de dados para uma sondagem.

x	Y	z	C0	C1	C01	C234
-83474,303	-95523,989	0,25	0	1	1	0
-83474,303	-95523,989	0,75	0	0,2	0,2	0,8
-83474,303	-95523,989	1,25	0	0	0	1
-83474,303	-95523,989	1,75	0	0	0	1
-83474,303	-95523,989	2,25	0	0	0	1
-83474,303	-95523,989	2,75	0	0	0	1
-83474,303	-95523,989	3,25	0	0	0	1

A coordenada Z refere-se ao ponto médio do troço, isto é 0,25m, 0,75m, 1,25m ... ou 3,25m. C0, C1, C01 e C234 referem-se às proporções de cada uma das classes de análise sensorial ou agrupamentos observadas no troço em análise (variável indicatriz). Os valores podem ser 0 ou 1 ou podem estar no intervalo 0 a 1. Zero ou um significa que a análise sensorial foi uniforme no troço; valores intermédios significam mistura. Por exemplo, na linha 2 da tabela, 0,2 e 0,8 significam que 20% do troço foi analisado como sendo de classe 1 e 80% como sendo de classe 2, 3 ou 4. Os valores têm assim o significado de probabilidade de pertença a uma classe.

As tabelas 3.5 e 3.6 mostram os estatísticos básicos das classes de reacção ao óleo e dos teores em TPH nos suportes finais de 0,5m utilizados para a fase de modelação geoestatística, respectivamente para todos os horizontes e por horizonte.

Tabela 3.5 – Estatísticos básicos das classes de reacção ao óleo e teores em TPH (ppm) por classe de análise sensorial até à profundidade máxima de 3,5 metros.

Classes de análise sensorial	Classes	Média indicatriz (proporção de uns)	TPH				
			Nº de amostras	Média	Mediana	Desvio Padrão	Coef. de assimetria
Reacção ao óleo	0	0,466	29	44,14	30,0	37,37	3,28
	1	0,237	50	374,4	117,5	469,32	1,24
	0+1	0,703	79	253,16	45,0	406,52	1,93
	2+3+4	0,297	80	7680,25	2925,0	13986,29	3,27
	Total	-	159	3990,06	995,0	10596,94	4,70

Tabela 3.6 – Estatísticos básicos das classes de reacção ao óleo e teores em TPH (ppm) por classe de análise sensorial e por intervalo de profundidade.

Classes	Intervalo de profundidade	TPH				
		Nº de amostras	Média	Mediana	Desvio Padrão	Coef. de assimetria
0	[0,00; 0,50m]	3	30,00	30,00	0,00	-
	[0,50; 1,00m]	4	30,00	30,00	0,00	-
	[1,00; 1,50m]	4	35,00	32,50	6,12	0,82
	[1,50; 2,00m]	7	57,14	35,00	50,49	1,99
	[2,00; 2,50m]	6	59,17	35,00	54,27	1,76
	[2,50; 3,00m]	5	35,00	35,00	5,48	0,91
	[3,00; 3,50m]	-	-	-	-	-
1	[0,00; 0,50m]	9	199,44	100,00	291,08	2,19
	[0,50; 1,00m]	8	233,13	117,50	301,37	1,92
	[1,00; 1,50m]	8	416,25	212,50	493,70	1,17
	[1,50; 2,00m]	9	493,33	315,00	513,99	0,74
	[2,00; 2,50m]	10	447,50	190,00	536,56	0,87
	[2,50; 3,00m]	5	354,00	65,00	509,02	1,48
	[3,00; 3,50m]	1	325,00	325,00	-	-
0+1	[0,00; 0,50m]	12	157,08	47,50	262,55	2,66
	[0,50; 1,00m]	12	165,42	65,00	264,04	2,54
	[1,00; 1,50m]	12	289,17	52,50	441,37	1,76
	[1,50; 2,00m]	16	302,50	52,50	443,33	1,59
	[2,00; 2,50m]	16	301,88	52,50	445,41	1,56
	[2,50; 3,00m]	10	194,50	40,00	421,67	2,63
	[3,00; 3,50m]	1	1045,00	1045,00	-	-
2+3+4	[0,00; 0,50m]	4	15925,00	9077,50	13797,23	1,12
	[0,50; 1,00m]	9	21088,33	8810,00	25381,50	1,16
	[1,00; 1,50m]	19	5737,37	2400,00	12540,59	3,85
	[1,50; 2,00m]	22	6055,91	2795,00	11902,48	3,85
	[2,00; 2,50m]	13	3495,39	2760,00	2413,09	0,66
	[2,50; 3,00m]	11	5205,91	2565,00	7298,44	2,35
	[3,00; 3,50m]	2	7990,00	7990,00	2070,00	-

4. MODELAÇÃO GEOESTATÍSTICA DOS TEORES EM TPH

A modelação geoestatística dos teores em TPH encontra-se subdividida em três sub-capítulos que se desenvolvem seguidamente: (1) modelação das classes de reacção ao óleo por krigagem da indicatriz; (2) co-estimação dos teores em TPH condicionado às classes de reacção ao óleo, fase de teste de validação cruzada destinado a escolher o melhor mecanismo de incorporação da informação secundária; (3) co-simulação dos teores em TPH condicionado às classes de reacção ao óleo, fase de simulação de imagens.

A modelação desenvolveu-se na malha de blocos apresentada no final do capítulo 3. Em profundidade foram caracterizados 7 horizontes, cada um relativo a 0,5 metros. Na modelação das classes de reacção ao óleo foram considerados dois cenários, respectivamente com 2 e 3 categorias: (1) classes 0+1 e 2+3+4; (2) 0, 1 e 2+3+4.

4.1. MODELAÇÃO DAS CLASSES DE REACÇÃO AO ÓLEO

4.1.1. VARIOGRAMAS MULTIFÁSICOS E AJUSTAMENTO DE MODELOS TEÓRICOS

Foram calculados variogramas experimentais, por classe de reacção ao óleo, e os correspondentes variogramas multifásicos, sendo que apenas estes últimos foram ajustados. Foram consideradas duas direcções: uma omnidireccional horizontal e outra vertical.

Na figura 4.1 apresentam-se os variogramas das classes de reacção ao óleo (classes 0, 1 e 2+3+4) na direcção omnidireccional horizontal e na figura 4.2 os variogramas das mesmas classes na direcção vertical. Refira-se que o variograma referente ao agrupamento das classes 2+3+4 é igual ao do seu complementar, ou seja, das classes 0+1.

As figuras 4.1 e 4.2 mostram a existência de uma estrutura nos variogramas de todas as classes consideradas, sendo esta mais evidente na direcção vertical. Calcularam-se então os variogramas multifásicos e ajustaram-se modelos teóricos (figura 4.3). No ajustamento foi considerada uma razão de anisotropia de 13,33 vezes entre a amplitude nas direcções horizontal e vertical.

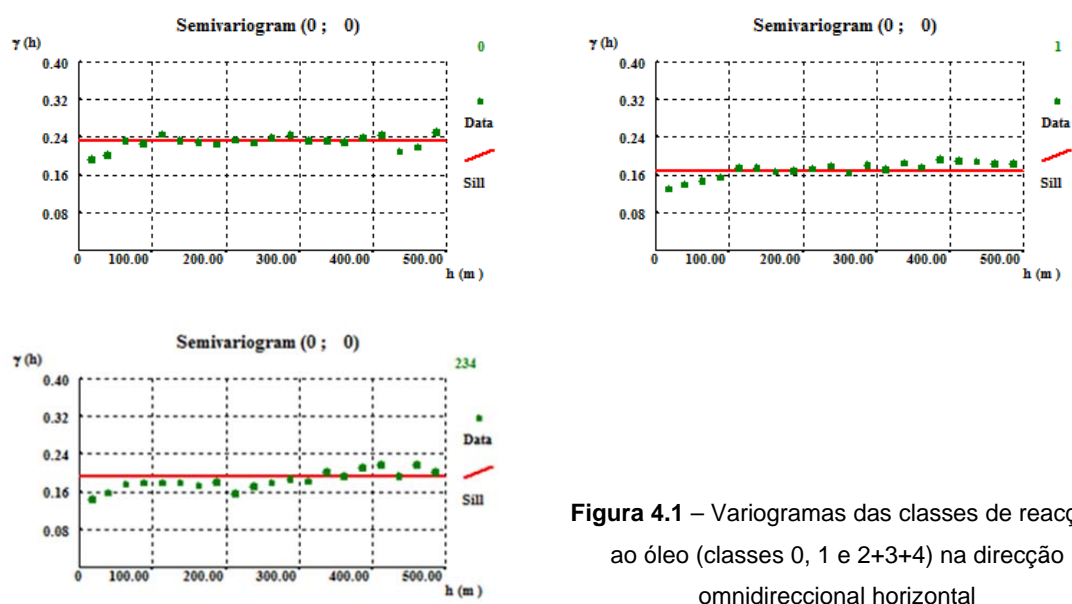


Figura 4.1 – Variogramas das classes de reacção ao óleo (classes 0, 1 e 2+3+4) na direcção omnidireccional horizontal

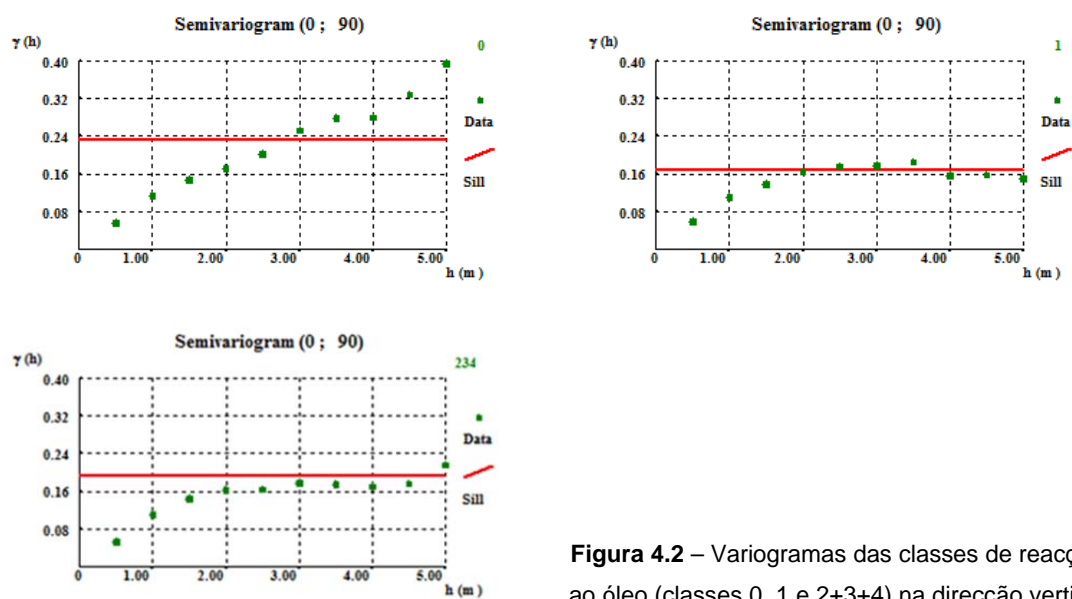


Figura 4.2 – Variogramas das classes de reacção ao óleo (classes 0, 1 e 2+3+4) na direcção vertical

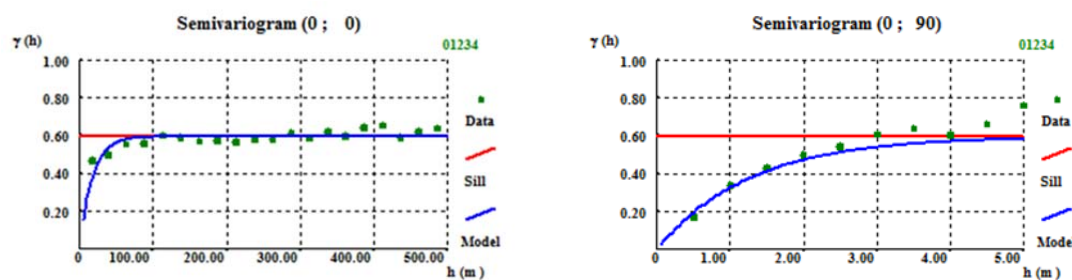


Figura 4.3 – Variogramas experimentais multifásicos das classes de reacção ao óleo nas direcções horizontal e vertical e ajustamento com modelo exponencial.

4.1.2. KRIGAGEM MULTIFÁSICA DAS CLASSES DE REACÇÃO AO ÓLEO

Nesta fase procedeu-se à estimação por krigagem multifásica das classes de reacção ao óleo (0, 1 e 234). Os resultados da krigagem são imagens de probabilidade de pertença de cada localização a cada classe, por horizonte (ver figuras 4.4 a 4.10).

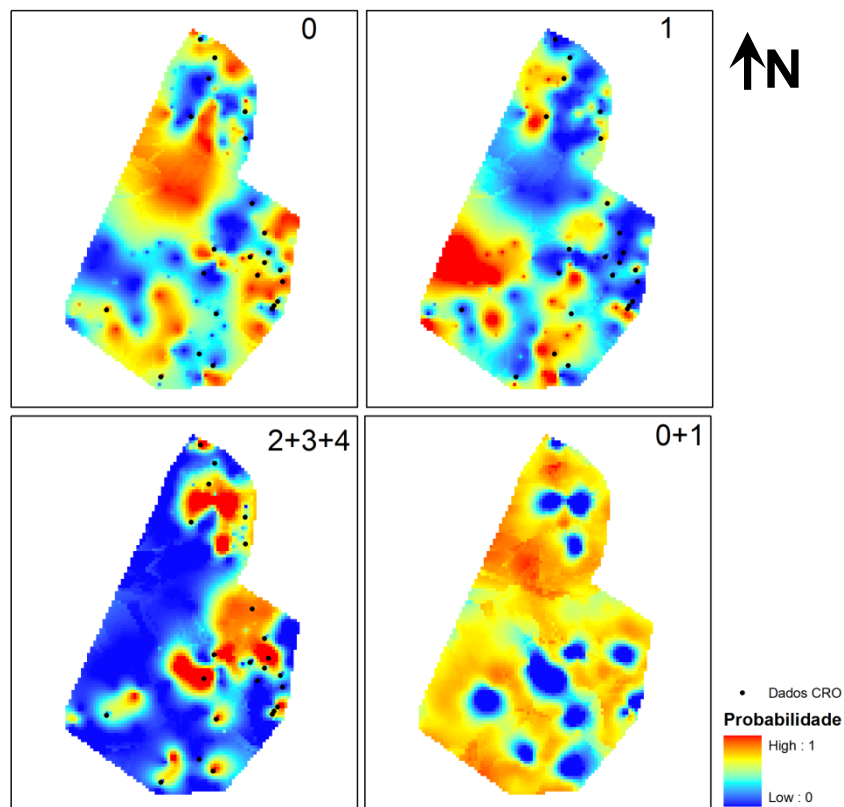


Figura 4.4 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 1 [0,00m; 0,50m].

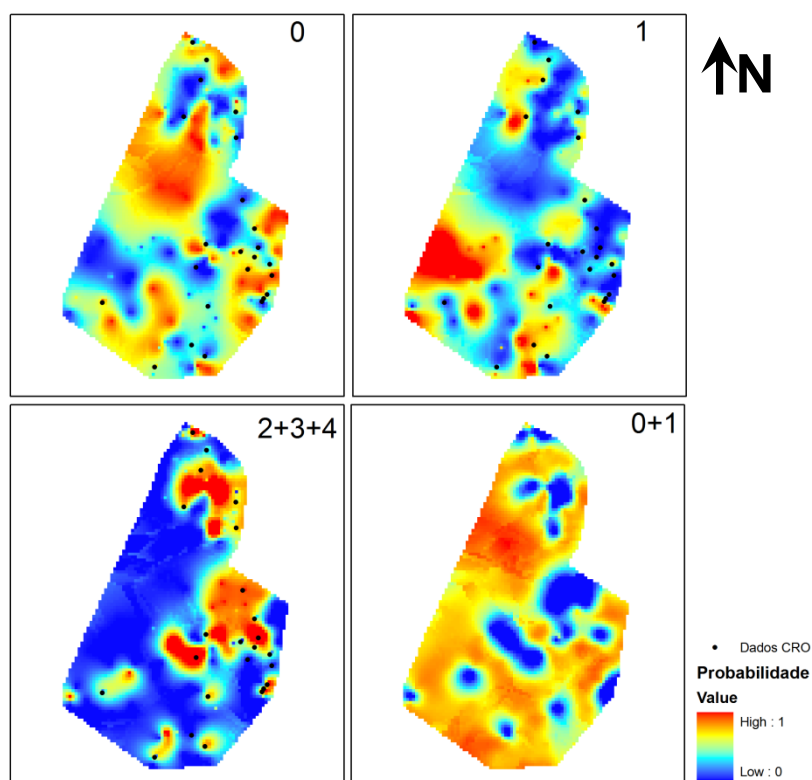


Figura 4.5 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 2 [0,50m; 1,00m].

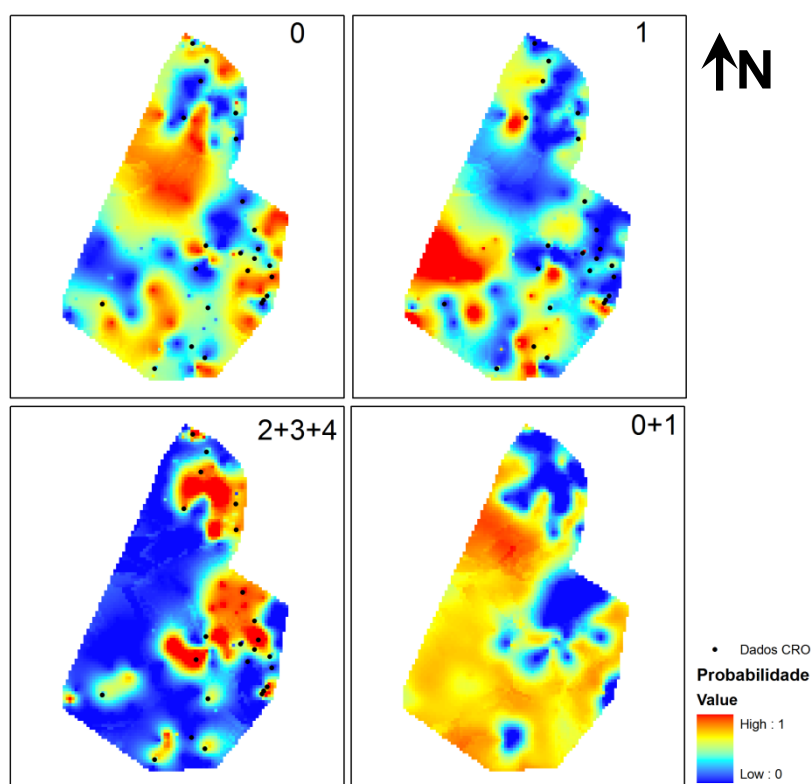


Figura 4.6 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 3 [1,00m; 1,50m].

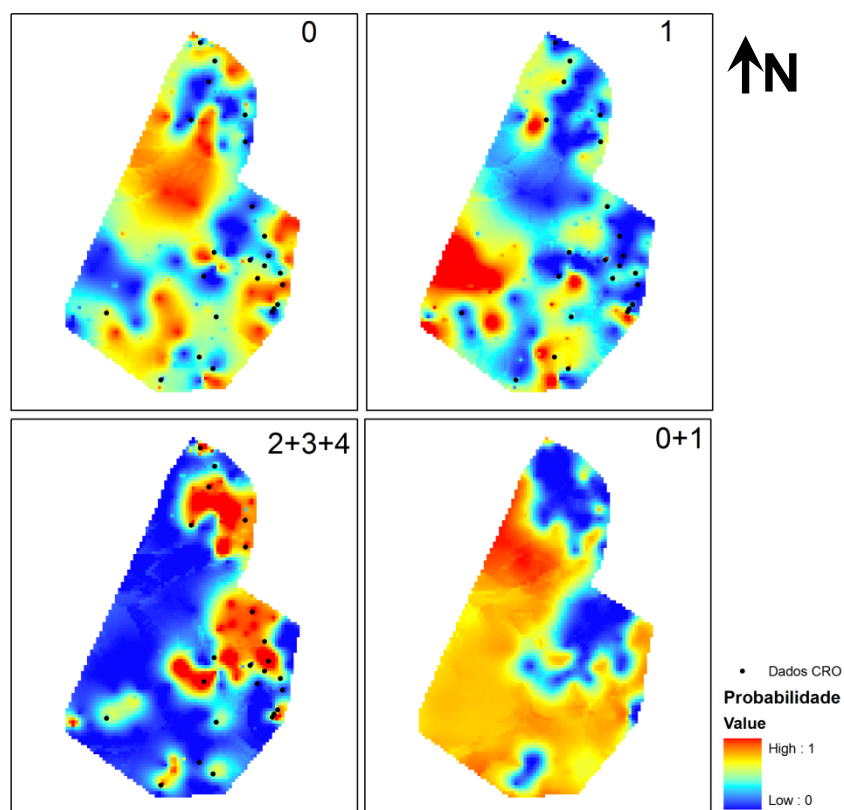


Figura 4.7 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 4 [1,50m; 2,00m].

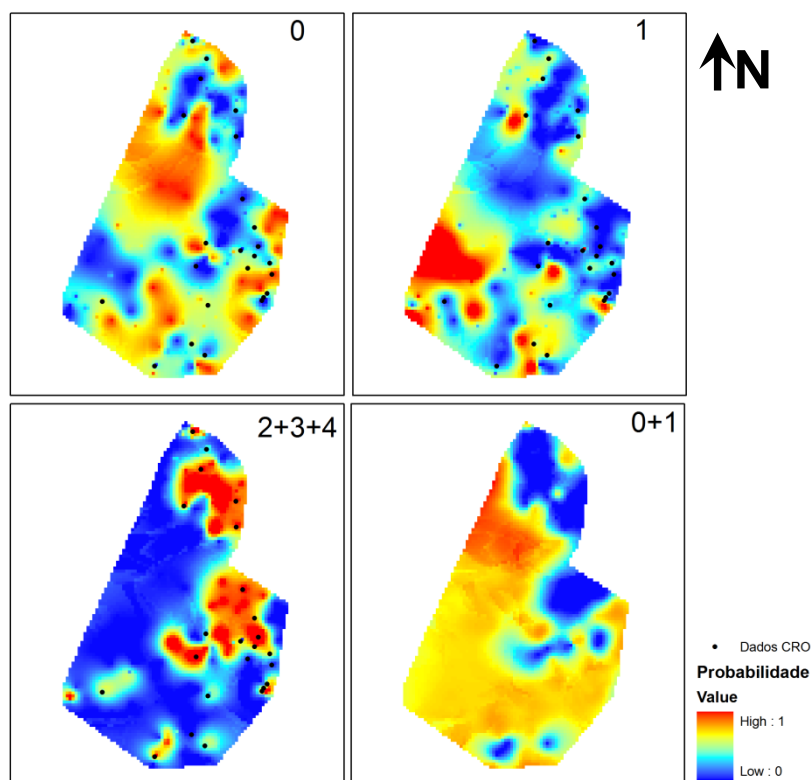


Figura 4.8 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 5 [2,00m; 2,50m].

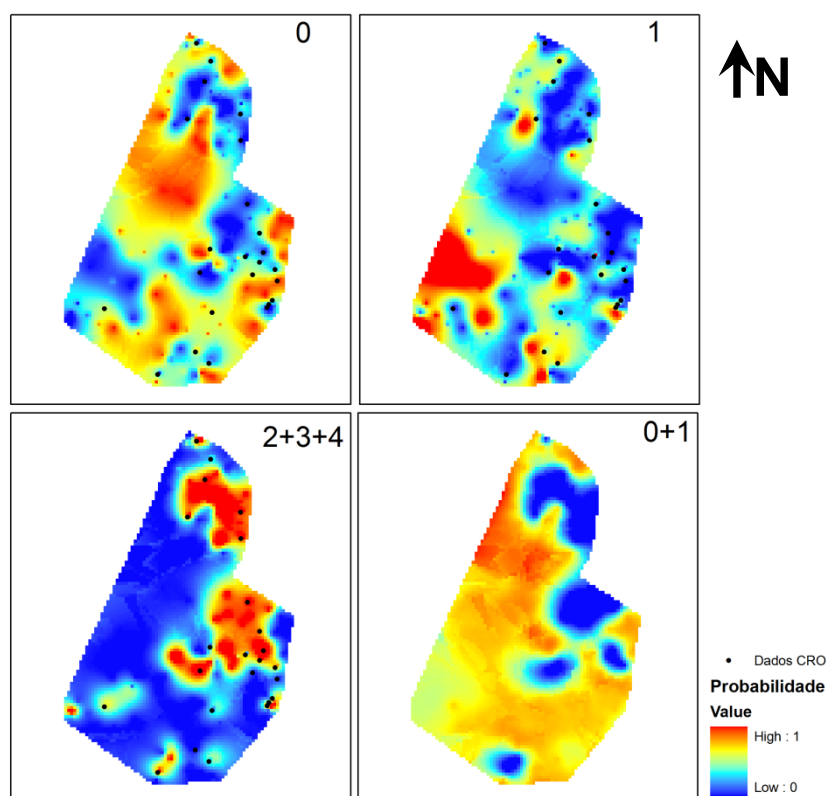


Figura 4.9 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 6 [2,50m; 3,00m].

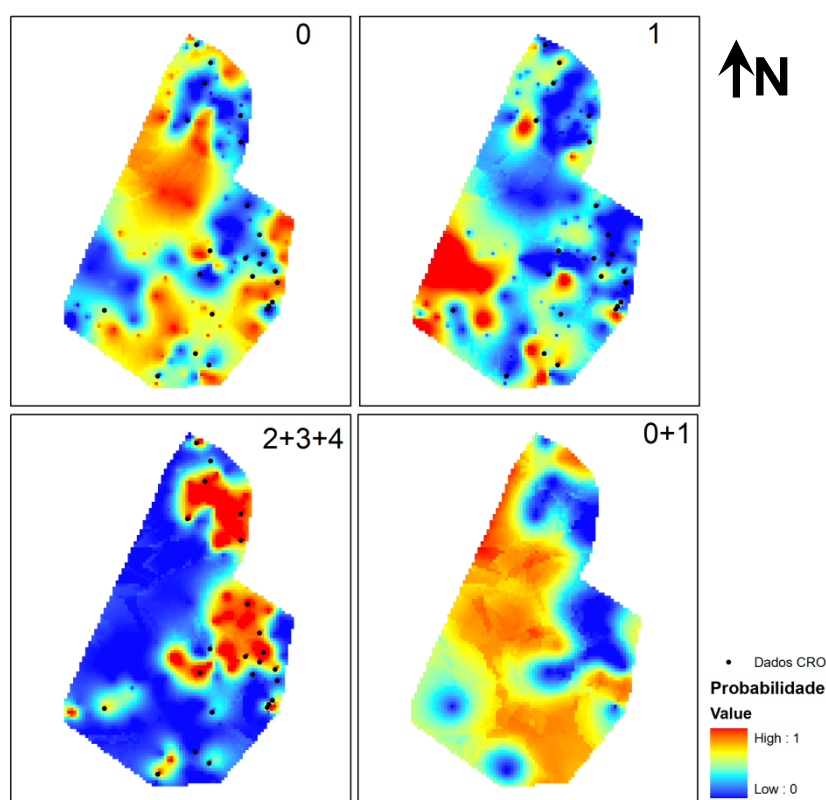


Figura 4.10 – Probabilidades estimadas por krigagem multifásica das classes de reacção ao óleo (0, 1, 2+3+4 e 0+1) para o horizonte 7 [3,00m; 3,50m].

Estas imagens de probabilidade foram transformadas para duas imagens finais das classes de reacção ao óleo mais prováveis em cada local. Os resultados para os 7 horizontes estão representados nas figuras 4.11 e 4.12.



Figura 4.11 – Classes de reacção ao óleo (0+1 e 2+3+4) para os 7 horizontes estimados.

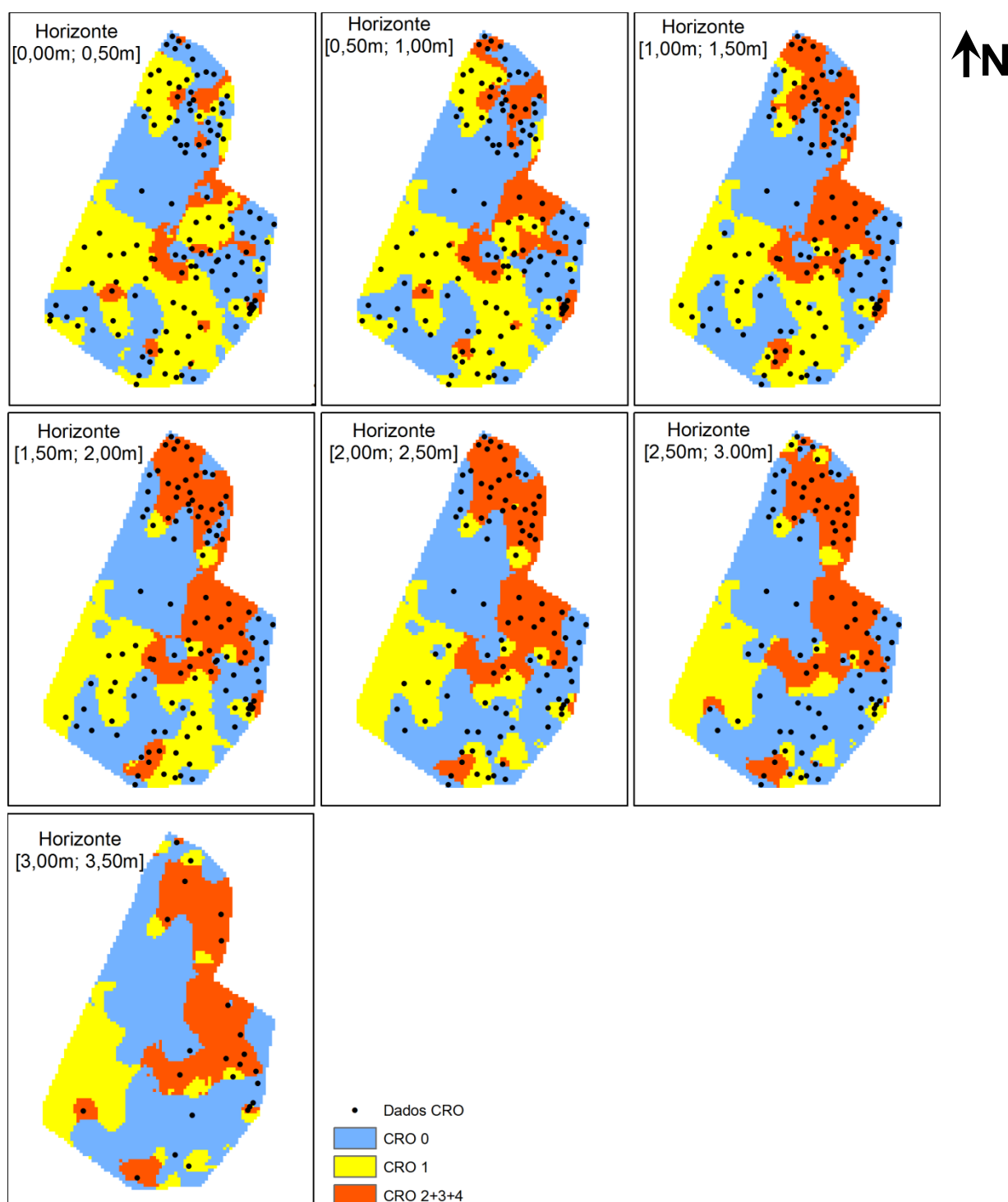


Figura 4.12 – Classes de reacção ao óleo (0, 1, 2+3+4) para os 7 horizontes estimados.

Esta sequência de imagens, dos níveis menos profundos para os níveis mais profundos, mostra que existe maior evidência de contaminação para profundidades superiores a 1m. Também o número de amostras disponíveis reduz-se muito rapidamente a partir do horizonte 6 (3m) mas é ainda visível a existência de amostras de classe 2+3+4, que delimitam as manchas

da respectiva classe. Todavia, este nível é estimado em grande parte à custa das amostras dos níveis superiores.

Tabela 4.1 – Proporções de cada classe por horizonte e proporções globais (imagem estimada / dados experimentais).

Horizonte	Amostras (#)	Classe 0	Classe 1	Classe 0+1	Classe 2+3+4
[0,00m; 0,50m]	131	0,473 / 0,49	0,430 / 0,37	0,903/0,86	0,097 / 0,14
[0,50m; 1,00m]	129	0,460 / 0,44	0,375 / 0,32	0,835/0,76	0,165 / 0,24
[1,00m; 1,50m]	123	0,447 / 0,41	0,321 / 0,23	0,768/0,64	0,232 / 0,36
[1,50m; 2,00m]	116	0,452 / 0,39	0,285 / 0,21	0,737/0,60	0,263 / 0,40
[2,00m; 2,50m]	98	0,481 / 0,41	0,253 / 0,20	0,734/0,61	0,266 / 0,38
[2,50m; 3,00m]	91	0,514 / 0,54	0,219 / 0,11	0,733/0,65	0,267 / 0,35
[3,00m; 3,50m]	24	0,522 / 0,25	0,202 / 0,17	0,724/0,42	0,276 / 0,58
[0,00m; 3,50m]	778	0,479 / 0,44	0,298 / 0,25	0,777/0,69	0,223 / 0,31

4.2. CO-ESTIMAÇÃO DOS TEORES EM TPH CONDICIONADO ÀS CLASSES DE REACÇÃO AO ÓLEO, FASE DE TESTE DE VALIDAÇÃO CRUZADA

4.2.1. CÁLCULO DE VARIOGRAMAS E AJUSTAMENTO DE MODELOS TEÓRICOS

Para os teores em TPH foram calculados variogramas experimentais nas direcções horizontal e vertical e ajustados modelos teóricos de tipo esférico sem efeito de pepita. Os resultados encontram-se representados na figura 4.13. Os variogramas aparecem em valores relativos até à unidade.

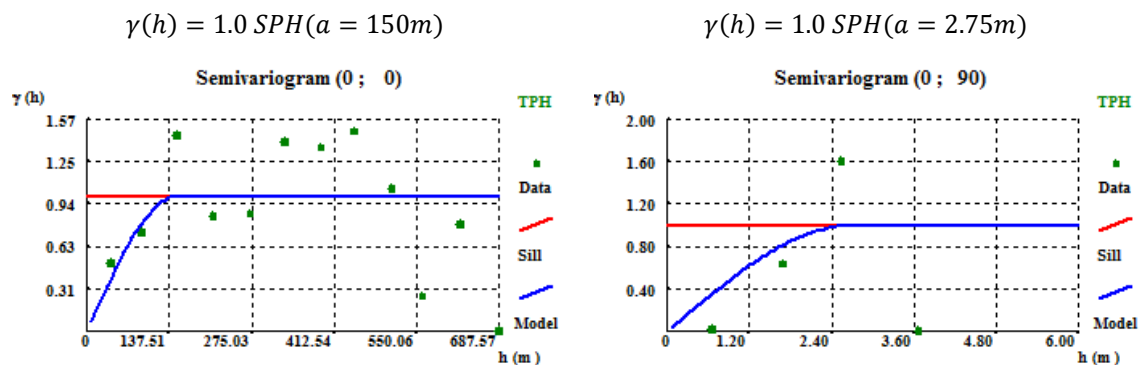


Figura 4.13 – Variogramas experimentais dos teores em TPH nas direcções horizontal e vertical e ajustamento com um modelo esférico.

4.2.2. TESTE DE VALIDAÇÃO CRUZADA E ANÁLISE DE RESULTADOS

Na simulação sequencial directa dos teores em TPH, para fazer a amostragem do histograma local, são estimados por krigagem simples a média e variância locais. Para aferir a melhor forma de incorporação das médias locais, foi conduzido um teste de validação cruzada. O teste consiste em comparar cada valor verdadeiro com os valores simulados segundo as várias variantes.

Neste teste foram seleccionadas aleatoriamente 20% das amostras (31 no total) para conjunto de controlo. Com as restantes 80%, foram simuladas 100 imagens dos teores em TPH da área de estudo. No final, para cada uma das 31 amostras, foram seleccionados os 100 valores simulados e analisados os desvios.

Os testes de simulação foram feitos com duas versões da SSD, respectivamente com um histograma global (versão clássica) e com histogramas locais, condicionados às classes de reacção ao óleo estimadas por krigagem da indicatriz. Na análise foram considerados os 2 cenários de estimação das classes de reacção ao óleo, com 2 e 3 conjuntos de classes. Os valores médios globais e condicionais por classe de reacção ao óleo e por horizonte estão listados nas tabelas 3.5 e 3.6 do capítulo 3. Em conjunto foram testadas sete variantes:

SSD_MG – simulação sequencial directa com médias globais (sem influência das classes de reacção ao óleo) – *cenário de referência*.

SSD_ML_CRO (01_234) e SSD_ML_CRO (0_1_234) – simulação sequencial directa com médias locais por classe de reacção ao óleo e condicionamento a histograma global. As classes de reacção ao óleo foram tratadas de duas formas: 2 conjuntos (0+1 e 2+3+4) e 3 conjuntos (0, 1 e 2+3+4).

SSD_HL_ML_CRO (01_234) e SSD_HL_ML_CRO (0_1_234) – simulação sequencial directa com médias locais por classe de reacção ao óleo e condicionamento aos histogramas locais. As classes de reacção ao óleo foram tratadas de duas formas: 2 conjuntos (0+1 e 2+3+4) e 3 conjuntos (0, 1 e 2+3+4).

SSD_HL_ML_CRO&Z (01_234) e SSD_HL_ML_CRO&Z (0_1_234) – simulação sequencial directa com médias locais por classe de reacção ao óleo e horizonte (profundidade) e condicionamento aos histogramas locais. As classes de reacção ao óleo foram tratadas de duas formas: 2 conjuntos (0+1 e 2+3+4) e 3 conjuntos (0, 1 e 2+3+4).

Nas figuras 4.14 e seguintes apresentam-se box-plots por amostra dos 100 valores obtidos com a simulação e indica-se também o valor verdadeiro da amostra (asterisco vermelho).

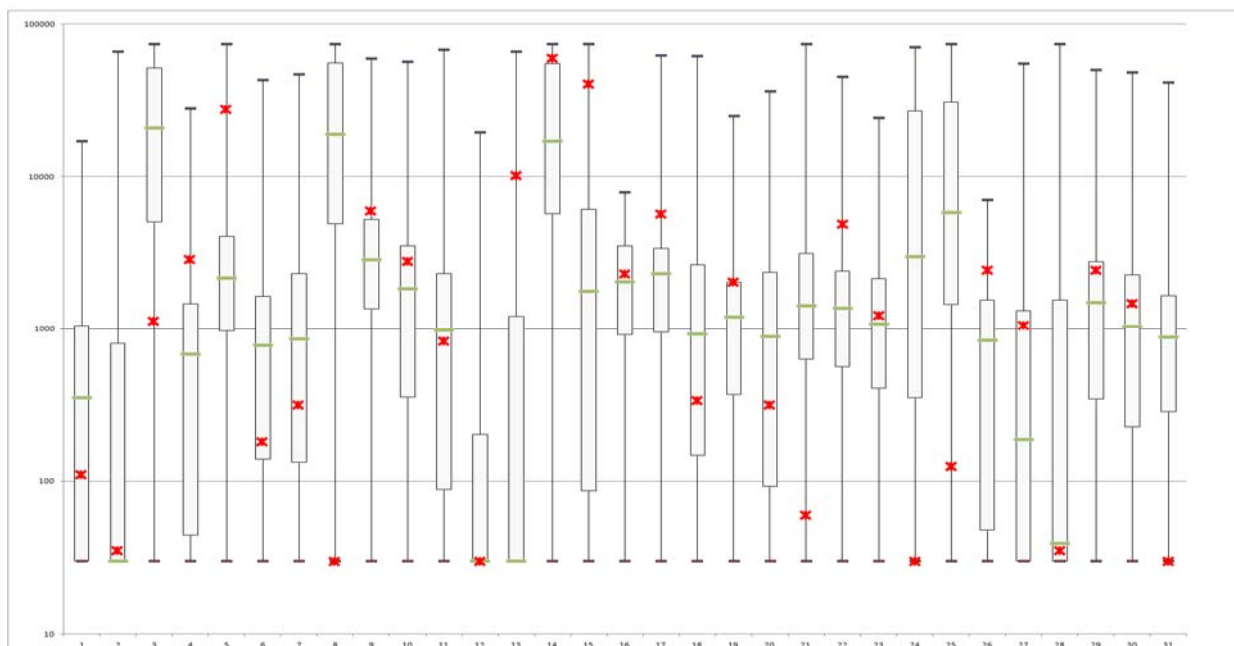


Figura 4.14 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base **SSD_MG**.

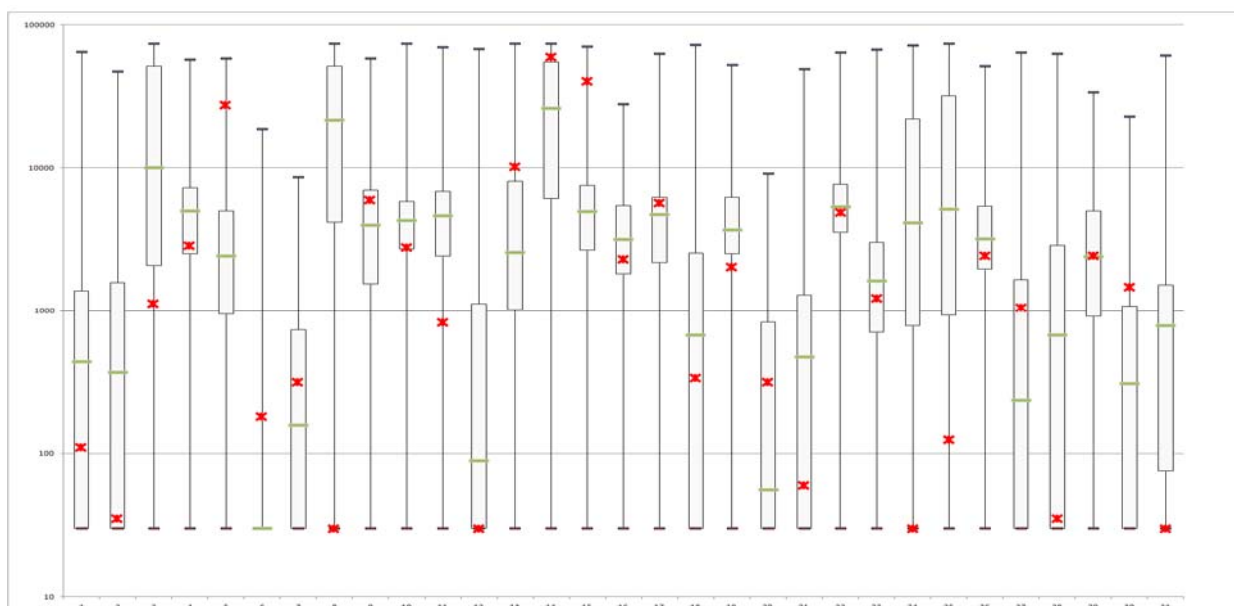


Figura 4.15 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base **SSD_ML_CRO (01_234)**.

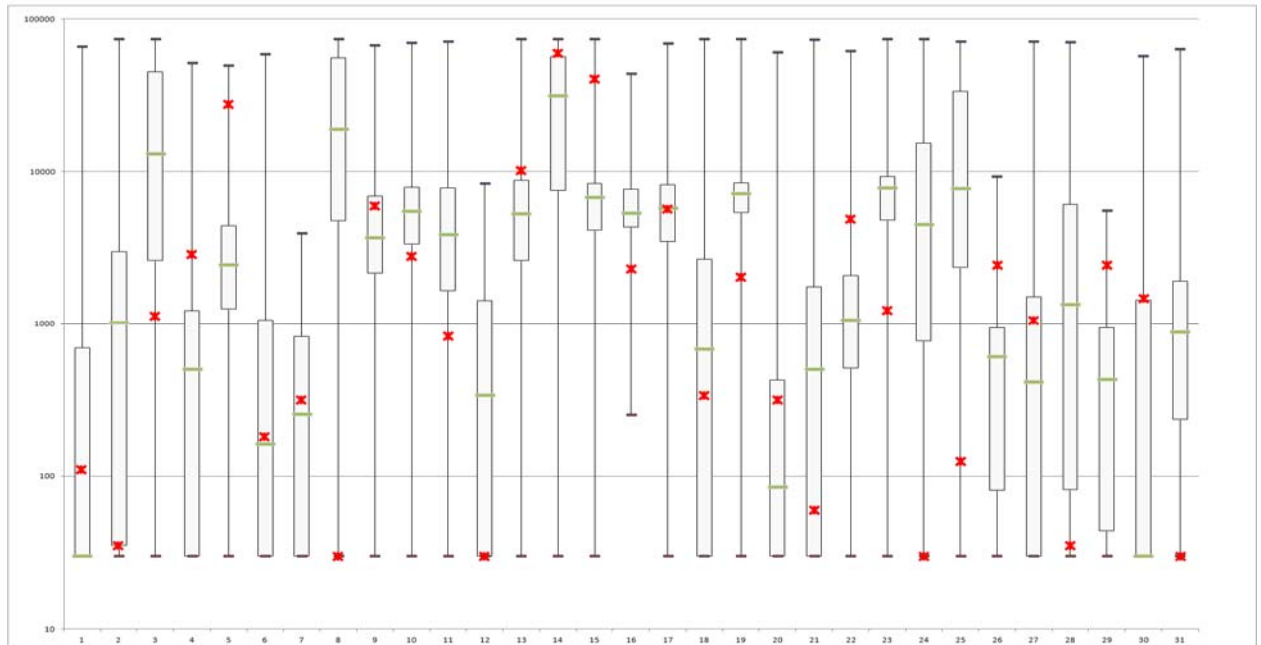


Figura 4.16 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base **SSD_ML_CRO (0_1_234)**.

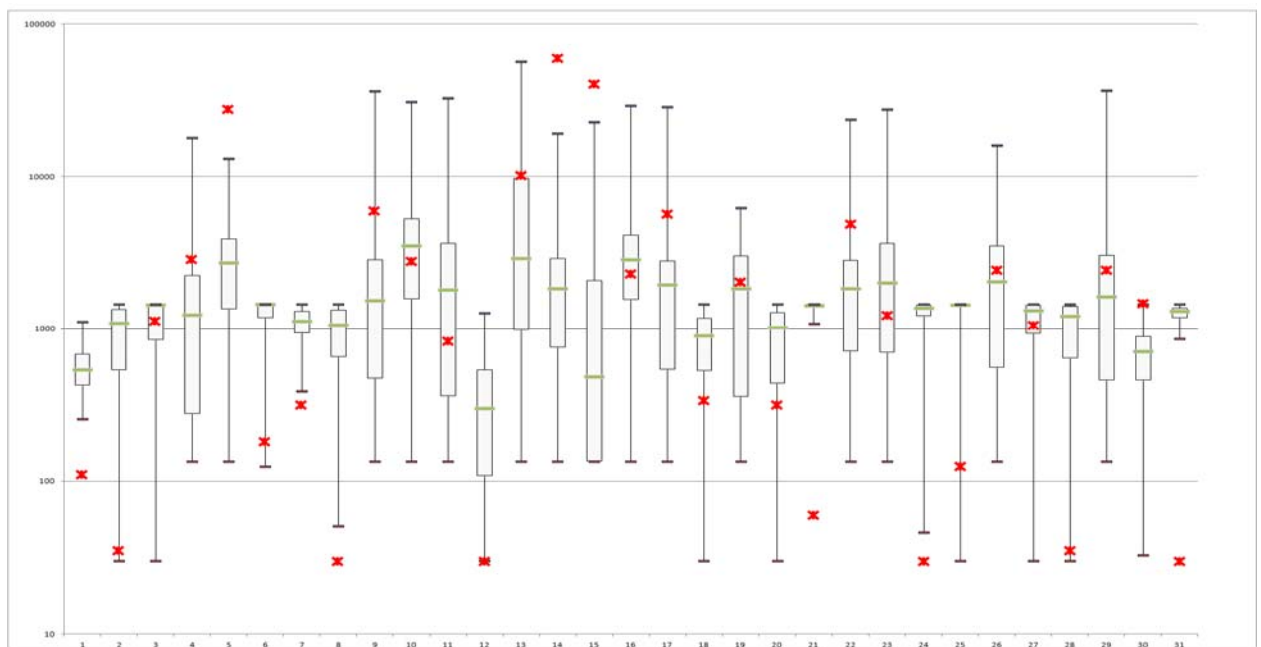


Figura 4.17 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base **SSD_HL_ML_CRO (01_234)**.

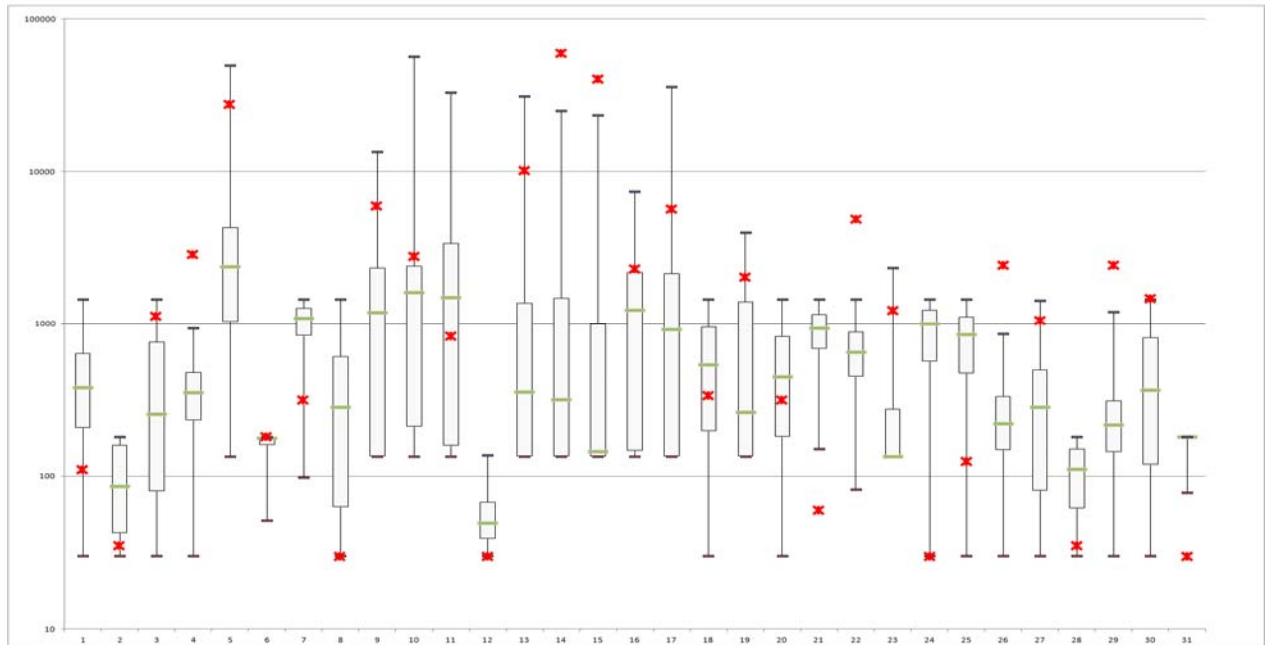


Figura 4.18 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base **SSD_HL_ML_CRO (0_1_234)**.

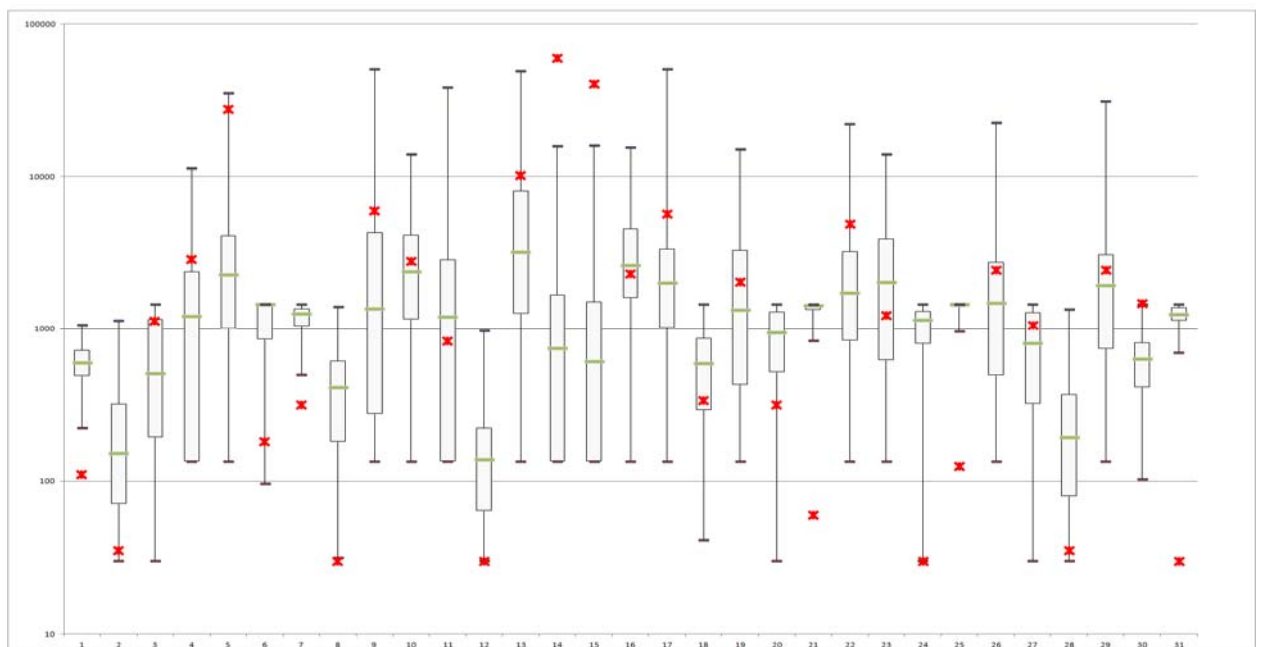


Figura 4.19 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base **SSD_HL_ML_CRO&Z (01_234)**.

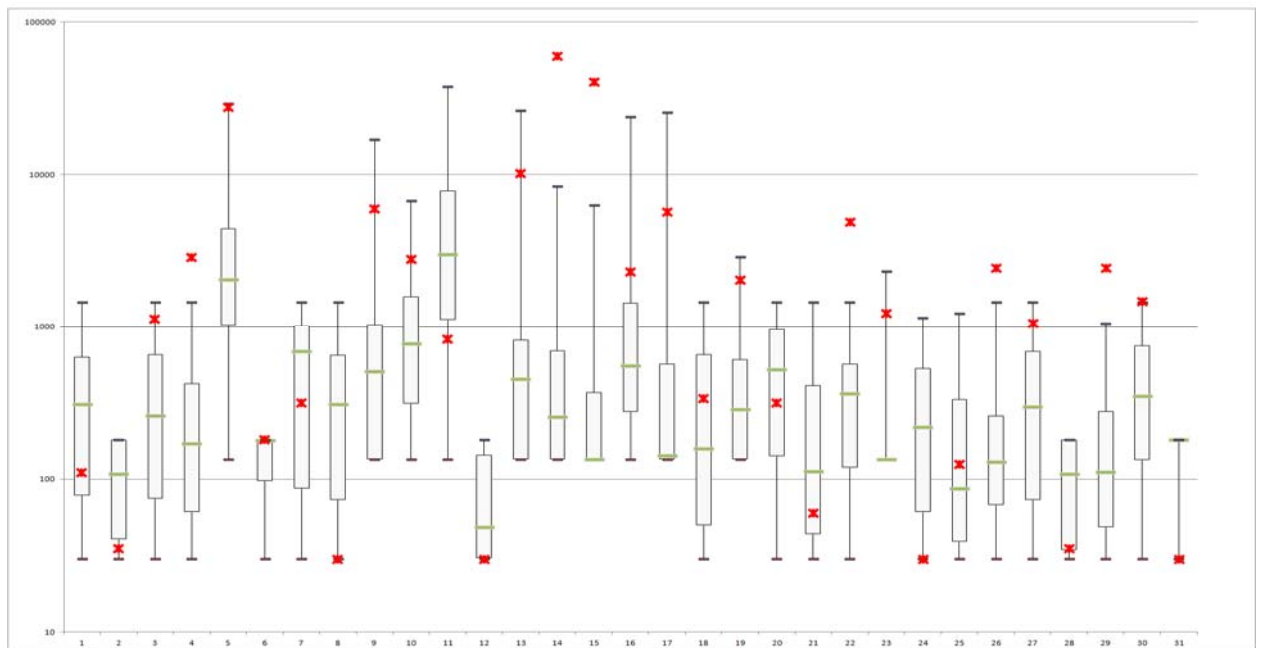


Figura 4.20 – Box-plot por amostra, representando o conjunto de valores simulados e sobreposição do valor verdadeiro da amostra (asterisco vermelho), cenário base **SSD_HL_ML_CRO&Z (0_1_234)**.

Na tabela 4.2 listam-se os resultados do teste de validação cruzada. Na tabela consideram-se as seguintes estatísticas:

EM – erro médio, médias dos desvios entre o valor verdadeiro e a média dos valores simulados.

EQM – erro quadrático médio, médias do quadrado dos desvios entre o valor verdadeiro e a média dos valores simulados.

Dispersão – média das variâncias dos valores simulados amostra a amostra. Mede a incerteza na caracterização dos teores, se conhecida a classe de reacção ao óleo.

BB; AA; AB e BA – dados relativos a um teste conduzido com um valor de corte de 1000 ppm. BB é o número de vezes em que o valor verdadeiro e a média dos valores simulados estão simultaneamente abaixo do teor de corte; AA idem ao anterior mas ambos acima. AB e BA significam más estimações, ou seja, contabilização do número de vezes que o valor verdadeiro e a média dos valores simulados não coincidem do ponto de vista de estarem acima ou abaixo do teor de corte.

% acertos – percentagem de amostras bem caracterizadas do ponto de vista de um teor de corte ou seja, $100 \cdot (AA+BB)/(AA+BB+AB+BA)$.

Tabela 4.2 – Análise estatística comparativa da diferença entre o valor verdadeiro e os 100 valores simulados amostra a amostra.

Método	EM	EQM	Dispersão	Teor de corte de 1000 ppm				
				BB	AA	AB	BA	%acertos
SSD_MG	-833	143,365,815	153,595,448	1	17		13	58
SSD_ML_CRO(01_234)	-2,187	136,838,356	179,056,462	3	17		11	65
SSD_HL_MG(01_234)	3,641	167,381,706	11,012,549	2	16	1	12	58
SSD_HL_ML_CRO(01_234)	3,787	169,874,247	11,390,565	8	14	3	6	71
SSD_HL_ML_CRO&Z(01_234)	3,701	170,048,518	13,390,593	7	14	3	7	68
SSD_ML_CRO(0_1_234)	-2,732	132,601,795	197,716,833	1	16	1	13	55
SSD_HL_MG(0_1_234)	4,582	174,486,362	5,795,421	12	7	10	2	58
SSD_HL_ML_CRO(0_1_234)	4,653	175,281,736	5,507,754	12	7	10	2	61
SSD_HL_ML_CRO&Z(0_1_234)	4,877	181,361,442	2,890,613	13	5	12	1	58

Estes resultados mostram que:

- Os valores absolutos do erro médio e do erro quadrático médio são muito elevados, o que mostra grande dispersão dos teores.
- As simulações sem condicionamento aos histogramas locais são sistematicamente sobre-estimadas, motivado por uma distribuição assimétrica e positiva dos dados experimentais e que é assumida para toda a área de estudo.
- Os erros quadráticos médios são todos da mesma ordem de grandeza.
- Os valores de dispersão têm 3 patamares, é sempre muito elevado quando se utiliza o histograma global, baixa muito quando se passa a utilizar um histograma local e mais ainda quando se consideram 3 modalidades de classes de reacção ao óleo. Este resultado vai ao encontro do que é esperado, porque se se considerarem histogramas locais a geração dos valores simulados por Monte Carlo é mais restritiva do ponto de vista do universo de soluções.
- No que respeita à análise com o valor de corte a 1000 ppm, os resultados obtidos com histogramas globais são pouco satisfatórios, por sobre-estimação sistemática. Os melhores resultados são obtidos quando se consideram 2 modalidades de classes de reacção ao óleo (0+1 e 2+3+4) e entre estes com médias locais sem discriminação do horizonte (cerca de 71%).
- Quando se consideram histogramas locais, existem valores verdadeiros de algumas amostras que ficam de fora do intervalo de valores simulados (ver figuras 4.18 a 4.21). Isto acontece principalmente quando a classe de reacção ao óleo estimada não coincide com a do valor experimental e então, na fase de simulação, os teores são simulados por Monte Carlo sobre o histograma “errado”. Este facto poderá ser obviado

em trabalhos futuros, fazendo a mistura ponderada dos histogramas de acordo com as probabilidades estimadas de pertença a cada classe ao invés da imagem binária classificada. A título de exemplo, para a variante SSD_HL_ML_CRO(01_234), se considerarmos apenas os 22 pontos em 31 do total onde a estimação das classes de reacção ao óleo está certa, os valores EM e EQM são respectivamente -28,47 e 2024646, ou seja, os desvios quase não têm enviesamento e têm baixa magnitude média.

Tendo em conta os resultados obtidos, considerou-se esta metodologia de utilização dos histogramas locais como evidenciando potencial para melhorar os resultados da simulação e optou-se por fazer a simulação de imagens com a variante SSD_HL_ML_CRO(01_234), que é aquela que gera a melhor taxa de acertos quando se considera um valor de corte de 1000 ppm, ligeiramente superior a 70%.

Mesmo considerando que a informação secundária é mais densa do que a informação principal, mas não é contínua e necessita de ser estimada, este resultado em particular mostra que existem vantagens na utilização da informação secundária das classes de reacção ao óleo para o controlo da distribuição espacial dos teores.

4.3. CO-SIMULAÇÃO DE IMAGENS DOS TEORES EM TPH CONDICIONADO ÀS CLASSES DE REACÇÃO AO ÓLEO

4.3.1. APRESENTAÇÃO E VALIDAÇÃO DE RESULTADOS

Foram simuladas 100 imagens de teores de TPH condicionadas às médias locais por classe de reacção ao óleo (nas classes 0+1 e 2+3+4). Nas figuras 4.21 a 4.27 mostram-se imagens de três realizações para cada horizonte. Na figura 4.28 mostram-se os variogramas experimentais de uma imagem simulada e o modelo teórico ajustado aos dados experimentais. Na figura 4.29 mostra-se a imagem média e na figura 4.30 a variância local dos mapas simulados. Finalmente na figura 4.31 mostra-se o histograma cumulativo condicional dos teores em TPH de uma imagem simulada por classe de reacção ao óleo.

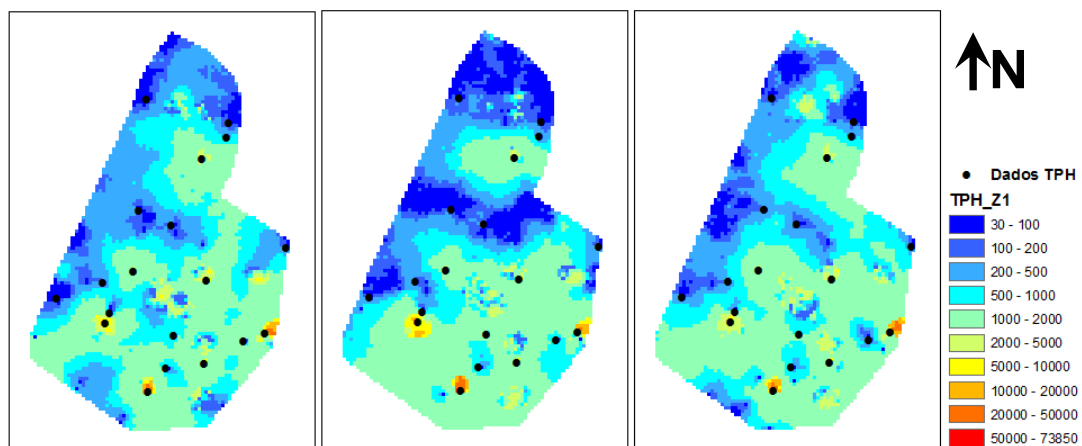


Figura 4.21 – Três imagens simuladas de TPH para o horizonte (0-0,5m).

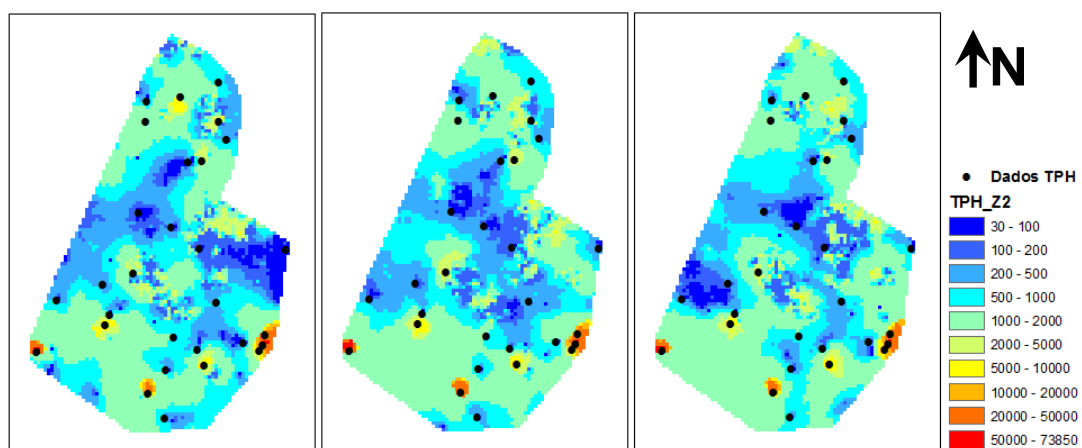


Figura 4.22 – Três imagens simuladas de TPH para o horizonte (0,5-1m).

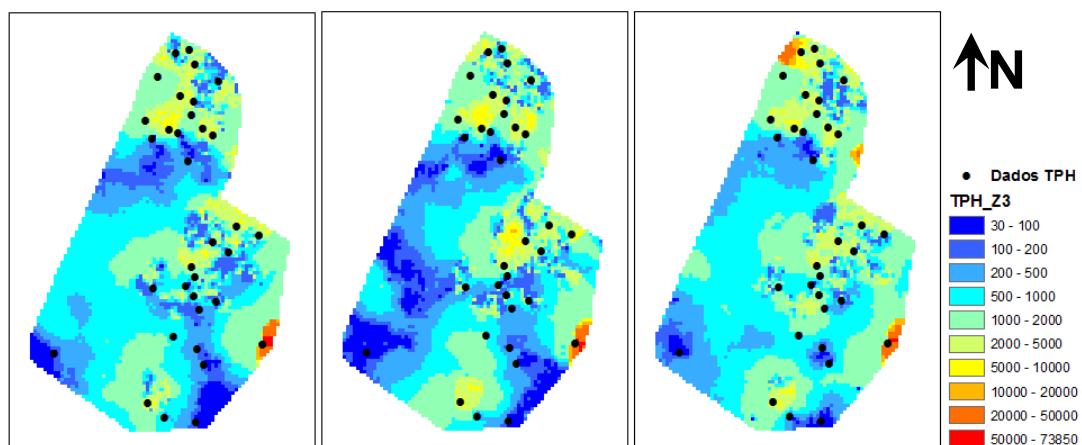


Figura 4.23 – Três imagens simuladas de TPH para o horizonte (1-1,5m).

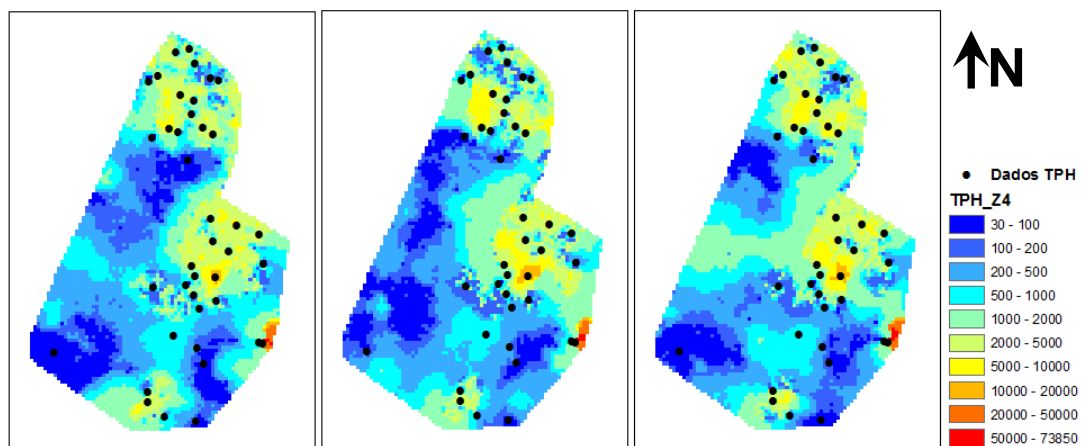


Figura 4.24 – Três imagens simuladas de TPH para o horizonte (1,5-2m).

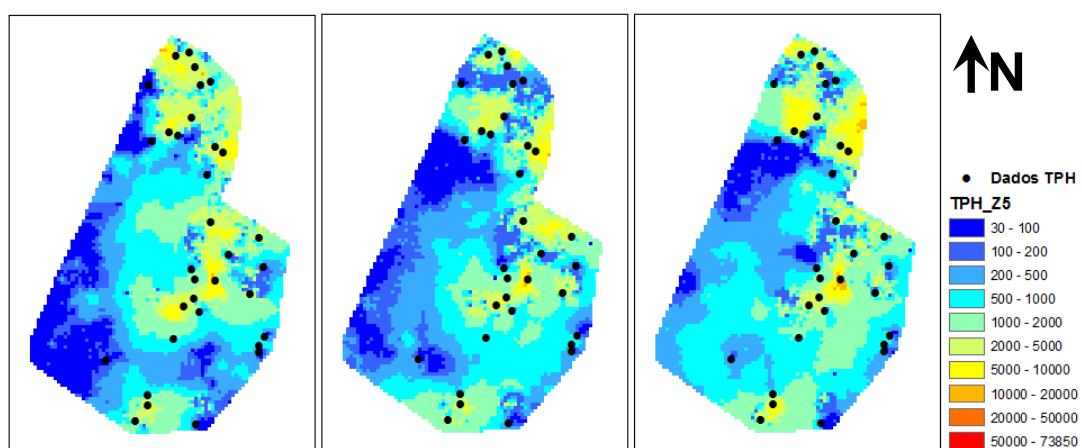


Figura 4.25 – Três imagens simuladas de TPH para o horizonte (2-2,5m).

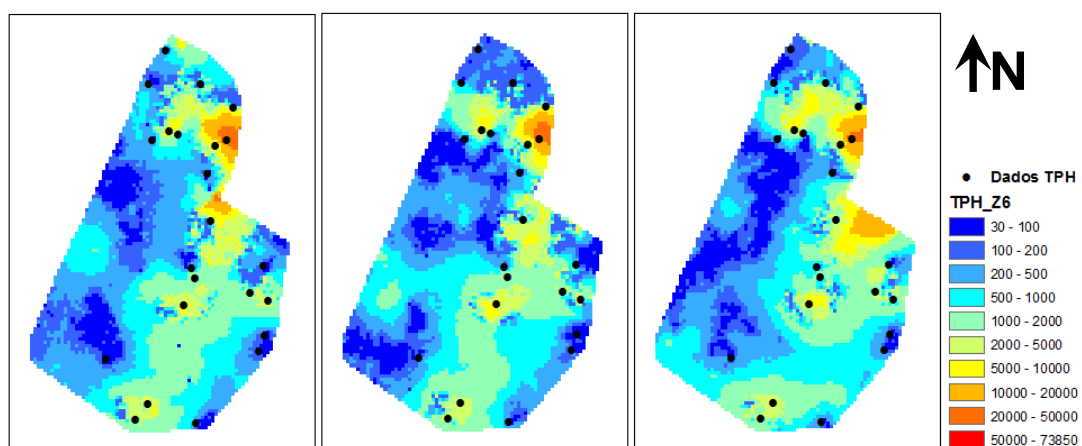


Figura 4.26 – Três imagens simuladas de TPH para o horizonte (2,5-3m).

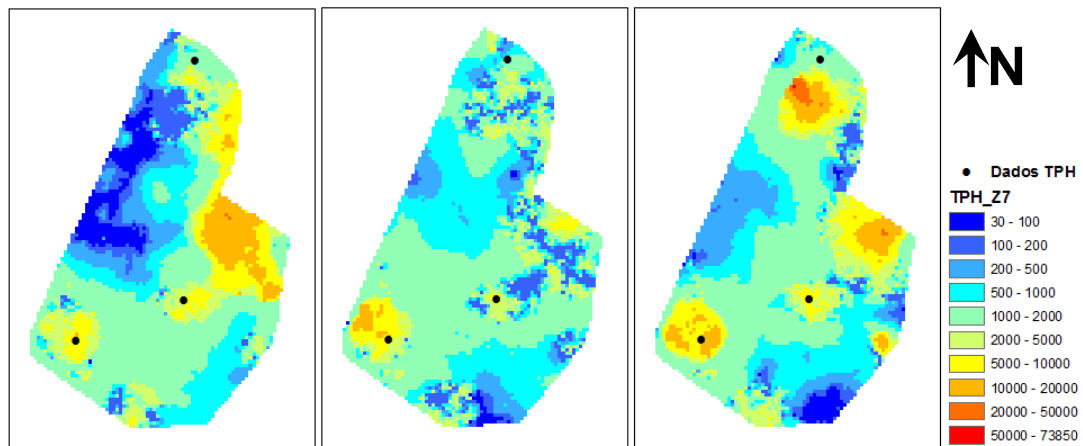


Figura 4.27 – Três imagens simuladas de TPH para o horizonte (3-3,5m).

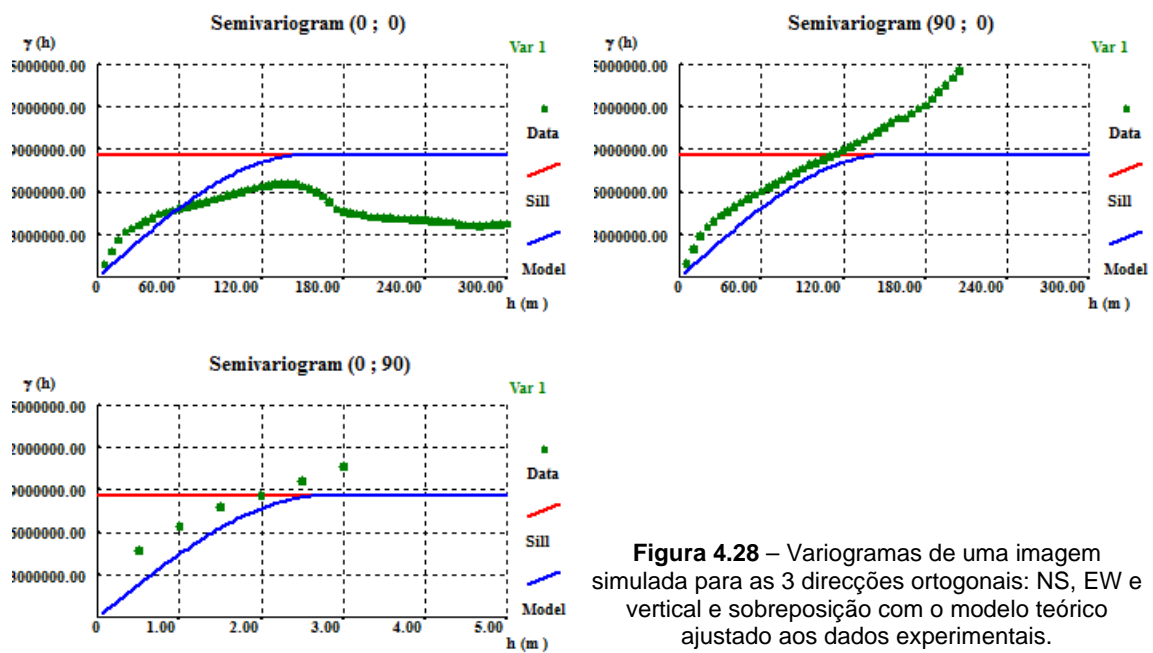


Figura 4.28 – Variogramas de uma imagem simulada para as 3 direcções ortogonais: NS, EW e vertical e sobreposição com o modelo teórico ajustado aos dados experimentais.

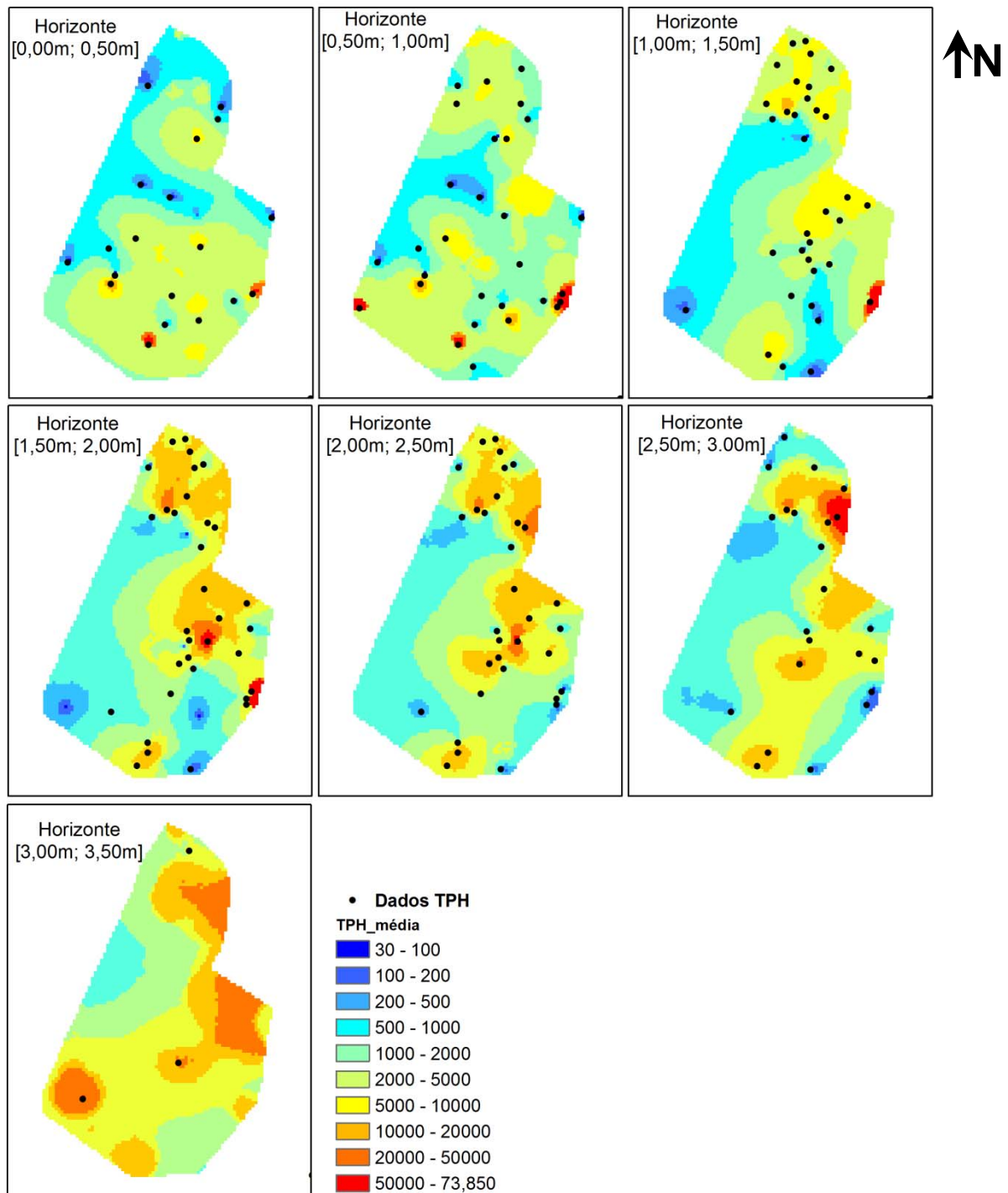


Figura 4.29 – Valores médios locais em TPH por horizonte calculados com base nas 100 simulações.

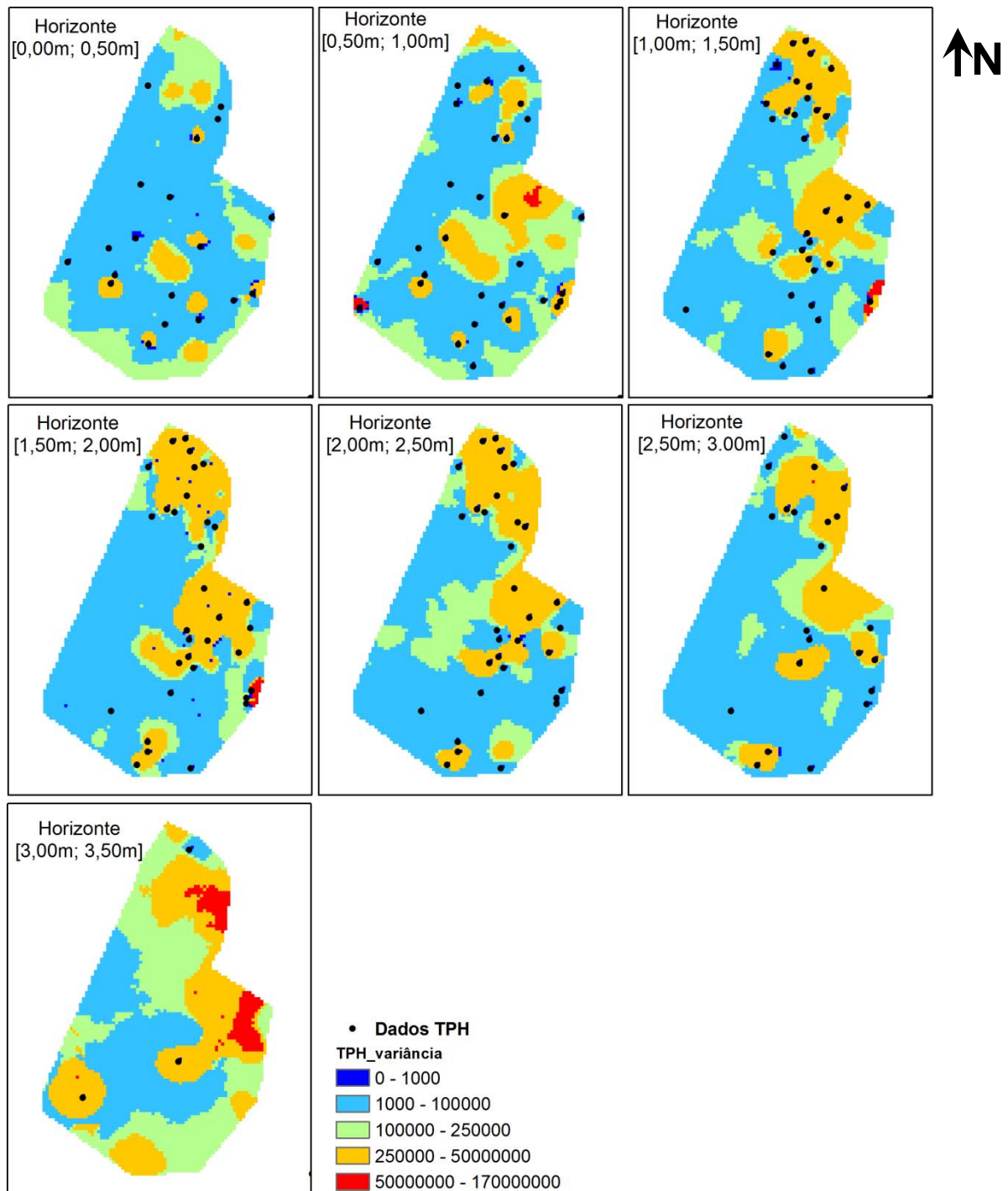


Figura 4.30 – Variâncias locais em TPH por horizonte calculados com base nas 100 simulações.

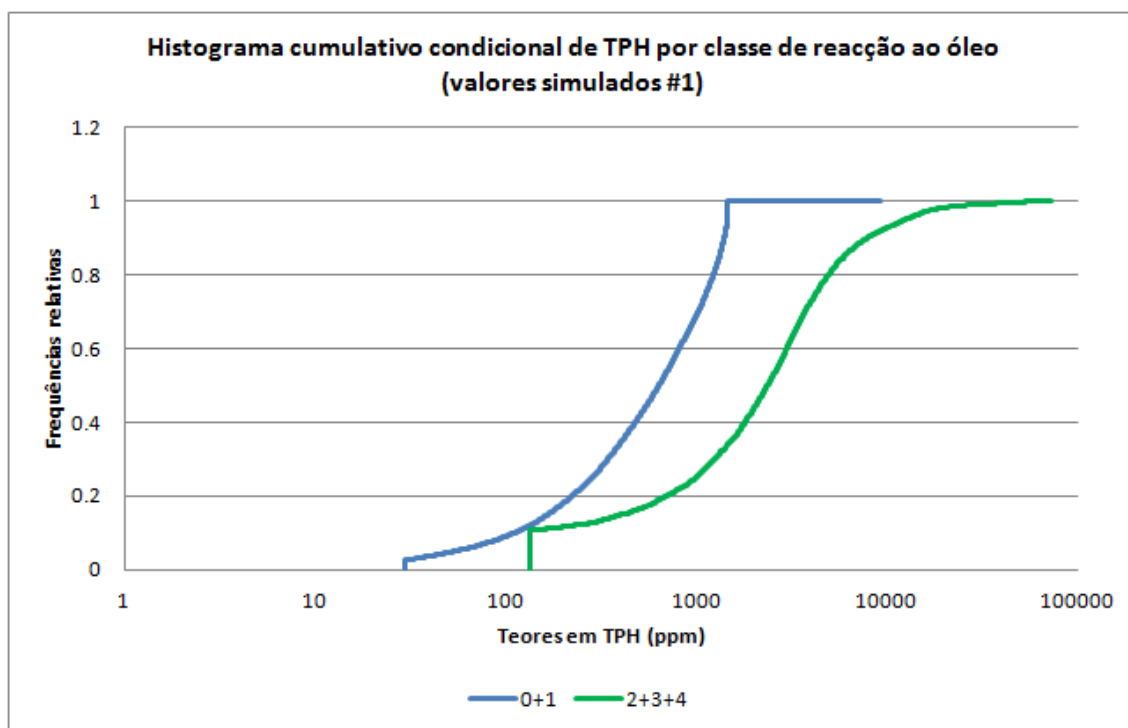


Figura 4.31 –Histogramas cumulativos condicionais de TPH por classe de reacção ao óleo para a realização #1.

Os resultados das simulações permitem retirar as seguintes evidências:

- 1) Observa-se zonamento dos valores simulados às classes de reacção ao óleo estimadas, de forma que os maiores valores pertencem à classe de reacção ao óleo 2+3+4 e os mais baixos à classe 0+1;
- 2) As transições entre valores baixos e valores altos é nalguns sítios gradual, noutros mais rápida, conforme o condicionamento imposto pela geometria das amostras em TPH;
- 3) Os três horizontes entre 1,5m e 3m apresentam as manchas mais extensas de teores elevados em TPH, e este facto é suportado por várias amostras em TPH. No horizonte seguinte (3-3,5m) também se observam manchas extensas de valores altos em TPH mas estas manchas resultam da extrapolação em profundidade do horizonte anterior dado existirem muito poucas amostras neste horizonte e seguintes. Desta forma a caracterização dos teores deve ser assumida como sendo válida até à profundidade de 3 metros.
- 4) Foram calculados variogramas para uma imagem simulada e observou-se correspondência com os modelos teóricos ajustados aos dados experimentais, pelo que é de esperar que todas as imagens reproduzam da mesma forma o modelo teórico ajustado aos dados experimentais.

- 5) A variância das imagens simuladas mostra valores iguais a zero sobre as amostras em TPH e valores próximo de zero na vizinhança destas, o que significa que as imagens são condicionais às amostras. Nos outros locais, a variância segue essencialmente um padrão de efeito proporcional, ou seja, apresenta valores mais baixos para teores médios mais baixos (classes de reacção ao óleo 0+1) e valores mais altos para teores médios mais altos (classes de reacção ao óleo 2+3+4).
- 6) Os histogramas cumulativos condicionais de TPH por classe de reacção ao óleo de uma imagem simulada mostram o zonamento dos teores em TPH por classe de reacção ao óleo, conforme o que é observado nos dados experimentais.

4.3.2. BINARIZAÇÃO DAS IMAGENS DOS TEORES EM TPH

A título de exemplo, as imagens simuladas foram binarizadas para um valor de corte (*threshold*) de 1000 ppm em TPH. Os resultados podem ser observados na figura 4.32 (probabilidade local de exceder o valor de 1000 ppm) e 4.33 (imagem binária). Estas imagens permitem determinar volumes de solo eventualmente contaminado, assim como identificar os respectivos locais de ocorrência.

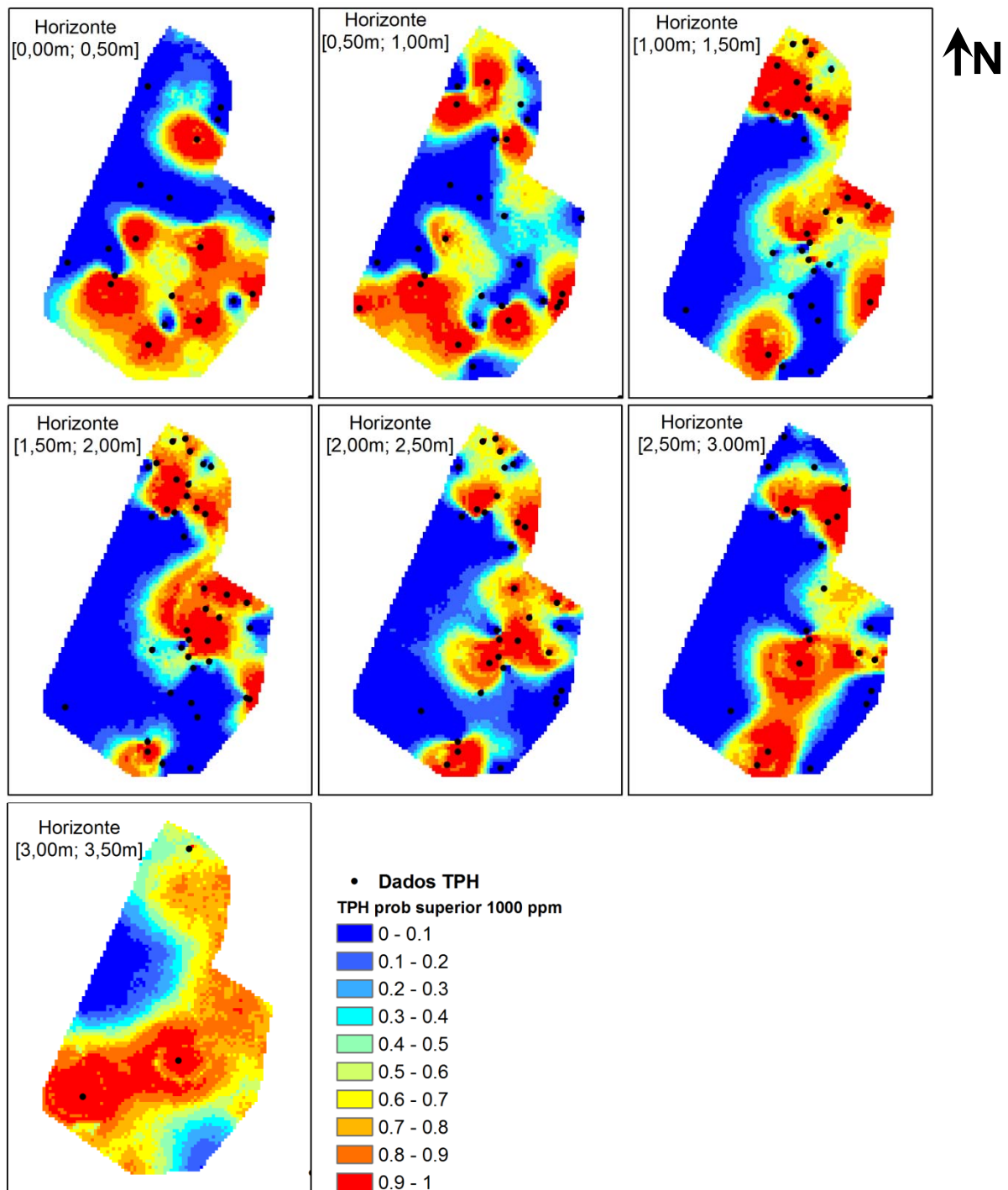


Figura 4.32 – Imagem de probabilidade local dos teores em TPH excederem o valor de corte de 1000 ppm para os 7 horizontes estimados.

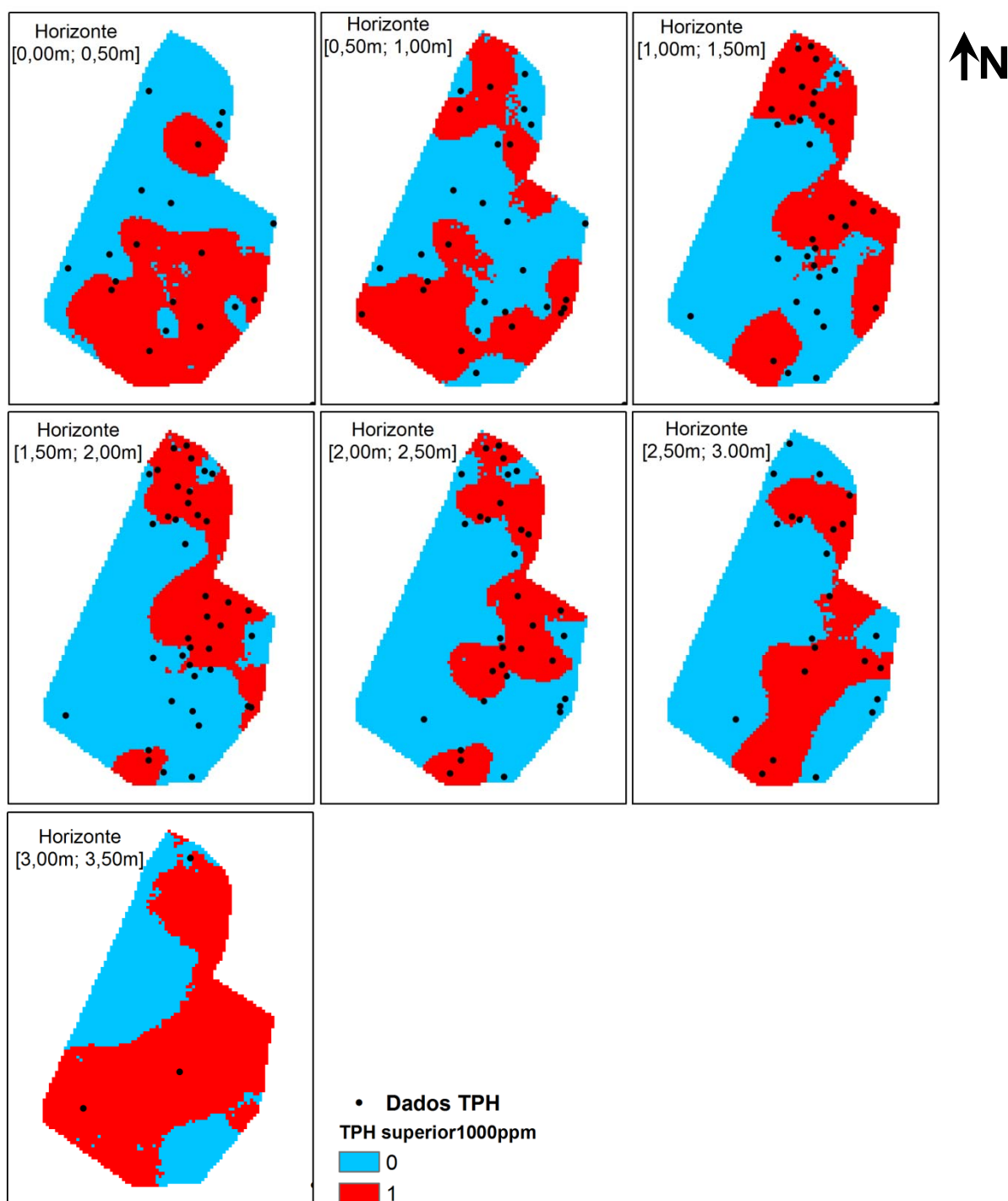


Figura 4.33 – Imagem binária de teores de TPH acima e abaixo de 1000 ppm, respectivamente categorias 1 e 0, para os 7 horizontes estimados.

5. CONCLUSÕES

O presente estudo permitiu caracterizar e delimitar áreas em que os solos apresentam valores elevados em TPH, com recurso a algoritmos de estimação e simulação geoestatística 3D e integração de informação principal e secundária. A simulação utilizou uma variante inovadora do algoritmo de simulação sequencial directa com condicionamento a histogramas e médias locais, o que permitiu reduzir localmente o grau de incerteza.

Da análise dos dados *soft* (informação secundária), através da estimação por krigagem simples, obteve-se uma sequência de imagens por classe de reacção ao óleo, onde foi possível constatar uma maior evidência de contaminação para profundidades superiores a 1m. Através da simulação foi igualmente possível verificar que o número de amostras disponíveis reduz-se muito rapidamente a partir dos 3m.

Posteriormente ao analisarmos a informação *hard* (informação primária) foi realizado um teste de validação cruzada, de modo a aferir qual a melhor forma de incorporação das médias locais. Assim, foram testados várias variantes da SSD, das quais a que apresentou melhores resultados foi a SSD_HL_ML_CRO_(01_234), ou seja, a simulação sequencial directa que recorre 2 modalidades de reacção ao óleo (0+1 e 2+3+4), condicionadas a histogramas e a médias locais. Esta variante da SSD apresenta 71% de acertos, ou seja, o nº de vezes que o valor verdadeiro corresponde ao valor simulado. Deste modo foi realizada a simulação dos teores de TPH tendo em conta a variante SSD_HL_ML_CRO_(01_234). Esta simulação permitiu verificar que existe um zonamento dos valores simulados às classes de reacção ao óleo estimadas, ou seja, os valores mais baixos simulados pertencem às classes de reacção ao óleo 0+1 estimadas e que os valores mais elevados pertencem às classes de reacção ao óleo 2+3+4. Os mapas simulados, e o respectivo mapa médio, mostraram que a variação da contaminação não é uniforme em profundidade, encontrando-se as manchas mais extensas com teores elevados de TPH nos horizontes entre 1,5 e 3m. A partir dos 3m, o nº de amostras é bastante reduzido, e os resultados obtidos resultam da extrapolação em profundidade do horizonte anterior. Desta forma a caracterização dos teores deve ser assumida como sendo válida até à profundidade dos 3 metros, uma vez que a partir desta profundidade não é possível validar os teores em TPH.

Este facto poderá estar associado a estarmos nas proximidades de uma linha de água e do nível freático se localizar próximo da superfície, pelo que a concentração de contaminantes em TPH se localiza na zona saturada do subsolo.

Através da binarização dos resultados é possível determinar volumes de solo eventualmente contaminados com teores de TPH acima e abaixo de 1000 ppm. Cerca de 44% da área de estudo apresenta teores acima dos 1000 ppm, ou seja, um volume de solos contaminados de aproximadamente 206.822 m³.

Recomenda-se que no futuro, sejam adoptadas técnicas de requalificação/ remediação dos locais que apresentam maiores valores de contaminação por TPH. Seja qual for a futura utilização do referido espaço, não existe um método de tratamento que seja melhor que outro. Assim recomenda-se que uma das medidas a adoptar poderia ser a da retirada dos solos contaminados com maiores concentrações em TPH, através de escavações e do transporte destes a aterros devidamente licenciados para o efeito, ou a adopção de medidas *in situ*, através do confinamento dos contaminantes.

6. BIBLIOGRAFIA

- Alexandre, J., Almeida, J. A., (1998) – geoMS – Geoestatistical Modelling Software. Centro de Modelização de Reservatórios Petrolíferos, Instituto Superior Técnico, Lisboa. On CDROM.
- Almeida, A., Soares, A., Reynaud, R., (1993) – Modelling the shape is Several Marble Types in a Quarry, Proc. of XXIV APCOM, *J. Elbrond, X. Tang*, Eds., 3, 452-459.
- Almeida, A., (1999) – Use of Geostatistical Model to Improve Reservoir Description and Flow Simulation in Heterogeneous Oil Fields, Tese de Doutoramento, *Instituto Superior Técnico*.
- Almeida, A. (2010) – Stochastic simulation methods for characterization of lithoclasses in carbonate reservoirs, 250-270 p.
- Almeida, C., Mendonça, J., Jesus, M., Gomes, A. (2000) – Sistemas Aquíferos de Portugal Continental. Instituto da Água e Centro de Geologia, 661pp;
- Brito, M.G. (2005) – Metodologia para a avaliação e remediação da contaminação por metais pesados em áreas industriais degradadas, 366 pp.
- Brito, M.G., Costa, C.N., Almeida, J.A., Vendas, D. & Verdial, P.H. (2006) - Characterization of maximum infiltration areas using GIS tools. *Engineering Geology*, (85) 14-18.
- DGMSG (1981) - Direcção Geral de Minas e Serviços Geológicos. Carta Geológica de Portugal na escala 1:50 000, Folha 34-B Loures, e respectiva Notícia Explicativa.
- DGGM (1988) – Direcção Geral de Geologia e Minas. Carta Neotectónica de Portugal Continental na escala 1/1 000 000, e respectiva notícia explicativa.
- Goovaerts, P., (1997) – Geostatistics for Natural Resources Evaluation, *Oxford University Press*, 483 p.
- Journel, A. G. e E. H. Isaaks (1984) – Conditional indicator simulation: Application to a Saskatchewan uranium deposit, *Mathematical Geology*, 16(7): 685-718.
- Journel, A.G., Alabert, F.G., (1989) – Non Gaussian data expansion in the earth sciences. *Terra Nova* 1, 123-134.
- Luís, A. A. G. (2004) – Utilização de informação auxiliar na estimação e simulação de variáveis regionalizadas. Tese de doutoramento não publicada. *Universidade Técnica de Lisboa, Instituto Superior Técnico*. 207 p.
- Matheron, G. (1963) – Principles of Geostatistics. *Econ. Geol.*, nº 58, 1246-1266.
- OMEE (2011) – Soil, Ground Water and Sediment Standards for Use Under Part XV.1 of the Environmental Protection Act. Ministry of the Environmental, April 15, 2011. PIBS#7382e01.

- Pereira, H. G. (1979) – Introdução às variáveis regionalizadas. Técnica 451-452, Lisboa, 89-96.
- Pereira, M.J., J. A. Almeida, M. G. Brito, A. Soares and E. Zungailia. 1997. Stochastic simulation of sediment quality for a dredging project. In V. Pawlowsky-Glahn, editor, Proceedings of IAMG'97. CIMNE, Barcelona, 2:899-904.
- Pereira, M. J., Almeida, J., Costa, C. & Soares, A. (2001) – Accounting for soft information in mapping soil contamination with TPH at an soli storage site. In Monastiez, P. Allard D. & Froidevaux, R. (eds.), geoENV III, Geoestatistic Environmental Applic., Kluwwer Academic Publish., Dordrecht, 475-486.
- Ruben & Almeida (2010) – Parallelization of sequential Gaussian, indicator and direct simulation algorithms, 1042-1052.
- Silva, T. & Almeida, J.A. (2009) – Cartografia digital da área impermeabilizada através de classificação de imagens LANDSAT 5 TM : um exemplo de análise espacio-temporal no concelho de Almada, Portugal. Revista Geociências, (28) Issue 3, 233-246.
- Soares, A., (1992) – Geoestatistcal estimation of multi-phase structures. Mathematical Geology 24 (2), 149-160.
- Soares, A., (2000) – Geoestatística para as Ciências da Terra e do Ambiente. Instituto Superior Técnico press, Lisboa, 206 pp.
- Soares, A., (2001) – Direct sequential simulation and cosimulation. Mathematical Geology 33 (8), 911-926.